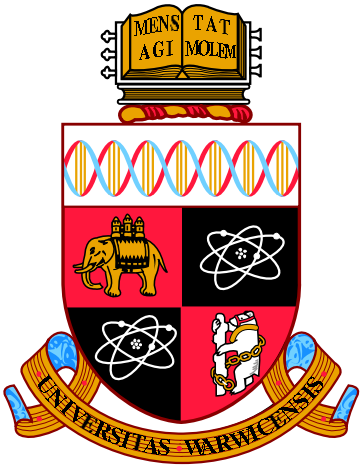


Computational Modelling of Lateral and Vertical Heterostructures of Transition Metal Dichalcogenides

by

Nelson Yeung

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Physics

September 2019

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisor Nicholas Hine for sharing the vast scientific and computational knowledge with me. Thank you for not giving up on me even though I have been very difficult to deal with.

I must also thank Joly Aarons, as I truly believe that without all his help and fruitful discussions, I would not have completed this thesis. Thank you Sam Brown for all the CSC help in my first year and being a great office pal. Thank you Sam Foster for being a great friend since undergraduate and basically walk this long journey of PhD with me with his own PhD.

My thoughts also go to Eduardo Mendive, Samuele Ferracin, Evaggelia Bisketzi, Dominic Branford, Jamie Friel, Francesco Francelli, Ju Ju Chang, Ken Kong, Kelvin Au, Brian Pak and Jasper Hau. Thank you very much for being such incredible friends and have provided me with immeasurable support mentally.

Finally, a heartfelt thank you to my parents for their love and support.

Declarations

This thesis is submitted to the University of Warwick as my application towards the degree of Doctor of Philosophy, and presents details of research carried out in the Theory Group of the Department of Physics between October 2015 and September 2019. The content of this thesis is my own work, unless stated otherwise, carried out under the supervision of Dr. N. D. M. Hine. No part of this thesis has previously been submitted for a research degree at any other institution.

Parts of this thesis have been published in the following paper:

1. Chapter 6: Producing Accurate Band Structures of Two-Dimensional Materials Using the GW Approximation

    - P. V. Nguyen, N. C. Teutsch, N. P. Wilson, J. Kahn, X. Xia, A. J. Graham, V. Kandyba, A. Giampietri, A. Barinov, G. C. Constantinescu, N. Yeung, N. D. M. Hine, X. Xu, D. H. Cobden, and N. R. Wilson, "Visualizing electrostatic gating effects in two-dimensional heterostructures", Nature **572**, 220–223 (2019)

Abstract

When certain materials in their bulk phase are reduced to a two-dimensional (2D) form, their electronic behaviour alters drastically, which has led to intense research into low-dimensional semiconductors within the condensed matter physics community. This thesis focuses on two relatively unknown heterostructures, lateral heterostructures (LHSs) and vertical heterostructures (VHSs), formed by one of the most studied and promising 2D materials, transition metal dichalcogenides (TMDC). LHSs comprises of two different 2D materials connected by covalent bond in-plane. VHSs are composed of different stacked layers that are weakly interacting.

The linear-scaling density functional theory (DFT) package ONETEP was used to study defects in small lattice mismatch and large lattice mismatch lateral heterostructures. We first examine the single sulphur vacancy formation energies across a LHS, then we present an approximation scheme to compute the critical thickness in which a dislocation will be formed for large lattice mismatch LHSs. The construction of a LHS for first-principles calculations is also presented.

We then move onto the work in collaboration with experimentalists, where angle resolved photoemission spectroscopy (ARPES) measurements with electrostatic gating was made for the first time on 2D materials to observe the band structure with the location of the conduction band edge. We show that the state of the art GW approximation is able to produce high quality band structure that is comparable to experiment. Here we have studied monolayer, bilayer and trilayer tungsten diselenide and tungsten disulfide. We also explore the possibility of using the virtual crystal approximation with GW to study the $Mo_{1-x}W_xS_2$ alloy.

Finally, we describe a novel scissor operator method for large scale DFT

calculations on VHSs. Due to the band gap problem within DFT, the DFT band alignment of certain TMDCs will not be realistic, and thus can predict incorrect band structures for certain VHSs. The present method corrects the band structure of a VHS using GW band alignment. We first describe the scissor operator implementation into ONETEP, then we show validation of this method with titanium disulphide–molybdenum disulfide VHS.

Abbreviations

ARPES **A**angle-**R**esolved **P**hoto**E**mission **S**pectroscopy

BZ **B**rillouin **Z**one

(C/V)BE (**C**onduction/**V**alence)-**B**and **E**dge

(E)DFT (**E**nsemble) **D**ensity **F**unctional **T**heory

EET **E**ffective **E**nergy **T**echnique

GF **G**reen's **F**unction

GGA **G**eneralised **G**radient **A**pproximation

GN **G**odby-**N**eeds

HF **H**artree-**F**ock

HOMO **H**ighest **O**ccupied **M**olecular **O**rbital

HL **H**ybertsen-**L**ouie

KS **K**ohn-**S**ham

LDA **L**ocal **D**ensity **A**pproximation

(L)DOS (**L**)ocal **D**ensity **O**f **S**tate

(L/V)HS (**L**ateral/**V**ertical) **H**etero**S**tructure

LNV **L**i, **N**unes and **V**anderbilt

MF **M**ean-**F**ield

NGWF **N**on-orthogonal **G**eneralised **W**annier **F**unction

ONETEP **O**rder-**N** **E**lectronic **T**otal **E**nergy **P**ackage

PAW **P**rojector **A**ugmented **W**ave

PBE **P**erdew-**B**urke-**E**rnzerhof

PL **P**hoto**L**uminescence

PPA **P**lasmon-**P**ole **A**pproximation

PW **P**lane **W**ave

QP **Q**uasi**P**article

RIM **R**andom **I**ntegration **M**ethod

RPA **R**andom **P**hase **A**pproximation

SCF **S**elf-**C**onsistent **F**ield

SPEM **S**canning **P**hoto**E**mission **M**icroscopy

SOC **S**pin **O**rbit **C**oupling

TMDC **T**ransition **M**etal **D**ichal**C**ogenide

VCA **V**irtual **C**rystal **A**pproximation

vdW **v**an **d**er **W**aals

XC E**X**change **C**orrelation ([functional])

Chapter 1

Introduction

Traditional silicon based technologies can only increase computing power by being made smaller and smaller. This poses a problem when the size reaches the quantum limit, where quantum effects can no longer be ignored, thus freezing technological advances. Therefore, research into new materials is of paramount importance.

Since the discovery that atomically thin or 2D materials can be made, indeed that they can be produced as simply as by pulling a scotch tape from a bulk crystal [2], research into these 2D materials has erupted in the science community. Initially, most attention was turned to graphene because it possesses many astounding properties, including high electron mobility and high mechanical resistance. However, despite it being semi-metallic, it lacked a band gap [3] which proved to be a difficult problem to address.

Therefore, in the recent years, attentions were turned to semiconducting inorganic layered materials, with the most notable class being transition metal dichalcogenides (TMDC). TMDC materials have proved to have remarkable applications [4, 5] in realms such as semiconductor electronics [6–8], spintronics [9–11] and optoelectronics [12–14]. This wide range of applications is further amplified by the fact that when these materials are made into atomic layers, they can be stacked vertically which are held together by the weak van-der-Waals force [15] to form vertical heterostructure (VHS), or connected seamlessly in-plane to form lateral heterostructure (LHS) [16], allowing the production of different materials containing unique characteristics.

This thesis focuses on using existing and new first-principles methods to model LHSs and VHSs. The next three chapters will begin with an introduction to the theoretical background used in this thesis: chapter 2 first details the theory behind density functional theory (DFT); chapter 3 then describes the shortcomings

of traditional DFT and the development of linear-scaling DFT, or more specifically, the linear-scaling DFT implementation within the ONETEP package is described; finally, chapter 4 discusses the theory behind the GW approximation, which addresses several problems within the DFT formalism.

In what follows are then three results chapters. In chapter 5, TMDC LHSs are studied. As using first-principles methods in LHSs requires several hundreds to several thousands, first-principles study of LHSs has been quite limited. A comprehensive understanding of structural defects is crucial, as having structural defects can be either detrimental or beneficial, depending on the targeted application. Here, we present the analysis of vacancy formation energies across LHSs with small lattice mismatch, then moving onto LHSs with large lattice mismatch, where we show an approximation method developed to compute the critical thickness of LHSs using first-principle calculations and provide a framework for building periodic LHSs with large lattice mismatch for DFT calculations.

In chapter 6, we then move onto the work in collaboration with experimentalists, where angle resolved photoemission spectroscopy (ARPES) measurements with electrostatic gating was made for the first time on 2D materials to observe the band structure with the location of the conduction band edge. We show that the state of the art GW approximation is able to produce high quality band structure that is comparable to the ARPES experiment. Here we have studied monolayer, bilayer and trilayer tungsten diselenide and tungsten disulfide. We also explore the possibility of using the virtual crystal approximation with GW to study the $Mo_{1-x}W_xS_2$ alloy. The purpose of this chapter is to provide the reader with a detailed recipe for using GW to provide experimentally comparable results.

In chapter 7, we describe a novel scissor operator method for large scale DFT calculations on VHSs. Due to the band gap problem within DFT, the DFT band alignment of certain TMDCs will not be realistic, and thus can predict incorrect band structures for certain VHSs. The present method corrects the band structure of a VHS using GW band alignment. We first describe the scissor operator implementation into ONETEP, then we show validation of this method with titanium disulphide–molybdenum disulfide VHS.

Finally, chapter 8 gives a summary of the thesis and describes potential avenues for future work.

Chapter 2

Ab-initio Electronic Structure Theory

2.1 The Many-Body Problem

Predicting certain properties of materials requires the consideration of all the interactions within the material. Naturally, the starting point is the Schrödinger equation,

$$i\hbar\frac{\partial}{\partial t}\left|\Psi(t)\right\rangle = \hat{H}\left|\Psi(t)\right\rangle, \tag{2.1}$$

which is the cornerstone of the study of quantum systems. It describes how a quantum state $|\Psi(t)\rangle$ is evolved over time t. Note the use of Dirac notation [17], where the wavefunction $\Psi(t)$ belongs in Hilbert space $\mathcal{H}$, or more formally $\Psi(t) \in \mathcal{H}$. $\hat{H}$ is the Hamiltonian operator, which characterises the total energy of the system. If the Hamiltonian is independent of t, the quantum system is said to be conservative and the above Schrödinger equation can be written as the so-called time-independent Schrödinger equation,

$$\hat{H}\left|\Psi\right\rangle = E\left|\Psi\right\rangle, \tag{2.2}$$

where E is the total energy of the system. We will restrict ourselves to such time-independent regimes throughout this and the next chapter of this thesis. In most cases, one is only concerned with atoms and molecules without time-independent interactions, so this restriction is indeed a valid one. In later chapters, when we describe excitations and photoemission, we will refer back to the time-dependent regimes as they are time-dependent phenomena.

In a many-body system, containing many electrons and nuclei, Ψ is called the many-body wavefunction, and it is a function of the positions of each electron

and nucleus. For example, if the position of the electrons and nuclei are $\mathbf{r}_n$ and $\mathbf{R}_m$, respectively, then we have

$$\Psi = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}; \mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_{N_I}). \tag{2.3}$$

In order to encapsulate the effects of all atomic interactions, one must include electron-electron, electron-ion and ion-ion interactions, which leads to the following form of the Hamiltonian:

$$\begin{aligned} \hat{H} = &-\frac{1}{2}\sum_{i=1}^{N_e}\boldsymbol{\nabla}_i^2 + \frac{1}{2}\sum_{i\neq j}^{N_e}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i=1}^{N_e}\sum_{\alpha=1}^{N_I}\frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} \\ &- \sum_{\alpha=1}^{N_I}\frac{1}{2M_\alpha}\boldsymbol{\nabla}_\alpha^2 + \frac{1}{2}\sum_{\alpha\neq\beta}^{N_I}\frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}. \end{aligned} \tag{2.4}$$

This is the many-body Hamiltonian, which defines the many-body problem. The Hartree system of atomic units is used here, where length, time, and mass are redefined such that $\hbar = e = m_e = 4\pi\epsilon_0 = 1$. This unit system will be employed throughout this thesis unless otherwise stated to simplify the equations.

Solving the Schrödinger equation with the full many-body Hamiltonian, eq. (2.4), is intractable for any real system, because the computational effort needed scales exponentially with the number of atoms. Efforts have gone onto reducing the computational time needed to solve the Hamiltonian. An obvious approximation to apply here is by taking into account that the timescale of motion of whole atomic nuclei is much longer than that associated with the movement of individual electrons. Hence we can assume that the motion of atomic nuclei and electrons can be separated. This is known as the Born-Oppenheimer approximation [18]. In addition, due to the large mass of the nucleus, it can be seen that the nuclear kinetic term in eq. (2.4) is negligible and that the nuclear interactions can be written as a simple parametric constant dependent only on the atomic positions of the system. Consequently, one only needs to consider the electronic part of the Hamiltonian, therefore the Hamiltonian can be reduced to

$$\hat{H} = \sum_{i=1}^{N_e}\left[-\frac{1}{2}\boldsymbol{\nabla}_i^2 + V(\mathbf{r}_i)\right] + \frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{2.5}$$

Although this Hamiltonian is marginally more tractable than the original Hamiltonian, it still scales exponentially with system size.

An early attempt to reduce the scaling of the problem was due to Hartree [19],

where the Coulomb term is replaced by an average potential, called the Hartree potential, from all the electrons:

$$V_{\mathrm{H}}(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' \,. \tag{2.6}$$

This is analogous the classical electrostatic potential, so that $n(\mathbf{r})$ is the electronic charge density. Due to the use of an average potential, this approach is sometimes considered to be a type of mean field approximation. This method is not accurate enough in many cases, due to the lack of exchange interactions. It was further extended by Fock [20] to include a non-local exchange potential that takes into the account of Pauli exclusion principle, where no two electrons can occupy the same quantum state. This extension, known as the Hartree-Fock approximation, gives qualitatively reasonable results for atoms. However, it fails for solids, because it neglects correlations and screening which are crucial for solids. Only after when Hohenberg and Kohn [21] proved perhaps two of the most important theorems in condensed matter physics, which led to the development of density functional theory (DFT), have first-principles or *ab-initio* calculations of solids have become widely feasible and useful.

2.2 Hohenberg-Kohn Theorems

2.2.1 Theorem I

Theorem I suggests that the density $n(\mathbf{r})$ can be used as a basic variable, since it uniquely determines the external potential $V(\mathbf{r})$, except by a constant difference, and hence the total energy E.

Suppose that there are two different external potentials $V^{(1)}(\mathbf{r})$ and $V^{(2)}(\mathbf{r})$, differing by more than a constant, which determine the same ground state density $n_0(\mathbf{r})$. In addition, the two external potentials lead to two different Hamiltonians $\hat{H}^{(1)}$, $\hat{H}^{(2)}$, and wavefunctions $\Psi^{(1)}$, $\Psi^{(2)}$. Since $\Psi^{(2)}$ is not the ground state of $\hat{H}^{(1)}$, one can write the inequality

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle , \tag{2.7}$$

where the last term can be written as

$$\begin{aligned} \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle &= \left\langle \Psi^{(2)} \middle| \hat{H}^{(2)} \middle| \Psi^{(2)} \right\rangle + \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} - \hat{H}^{(2)} \middle| \Psi^{(2)} \right\rangle \\ &= E^{(2)} + 2 \int \Big[V^{(1)}(\mathbf{r}) - V^{(2)}(\mathbf{r}) \Big] n_0(\mathbf{r})\, d\mathbf{r} \,. \end{aligned} \tag{2.8}$$

Note that the factor of 2 is the inclusion of spin degeneracy, as we assume a closed-shell system. Combining the above equations, eq. (2.7) and eq. (2.8), we arrive at

$$E^{(1)} < E^{(2)} + \int \Big[V^{(2)}(\mathbf{r}) - V^{(1)}(\mathbf{r})\Big] n_0(\mathbf{r})\, d\mathbf{r}\,. \tag{2.9}$$

If the same procedure is applied to $E^{(2)}$, we will arrive at the same result with the superscripts (1) and (2) simply interchanged,

$$\begin{aligned} \left\langle \Psi^{(1)} \middle| \hat{H}^{(2)} \middle| \Psi^{(1)} \right\rangle &= \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle + \left\langle \Psi^{(1)} \middle| \hat{H}^{(2)} - \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle \\ &= E^{(1)} + 2 \int \Big[V^{(2)}(\mathbf{r}) - V^{(1)}(\mathbf{r})\Big] n_0(\mathbf{r})\, d\mathbf{r} \end{aligned} \tag{2.10}$$

and

$$E^{(2)} < E^{(1)} + \int \Big[V^{(1)}(\mathbf{r}) - V^{(2)}(\mathbf{r})\Big] n_0(\mathbf{r})\, d\mathbf{r}\,. \tag{2.11}$$

By adding together eq. (2.9) and eq. (2.11), one will find a contradictory inequality,

$$E^{(1)} + E^{(2)} < E^{(1)} + E^{(2)}. \tag{2.12}$$

Therefore the initial condition that there are two different external potentials differing by more than a constant that produce the same non-degenerate ground state density has to be false. This establishes the desired result where the electronic density must uniquely determine the external potential to within a constant. Furthermore, since the Hamiltonian is completely determined by the charge density, the total energy can be considered as a functional of the charge density,

$$E[n] = T[n] + E_{\text{ext}}[n] + E_{\text{ee}}[n], \tag{2.13}$$

where T, E_{ext} and E_{ee} are the kinetic energy, external interaction and electron-electron interaction functionals, respectively. Therefore it can be said that with only the ground state density, all the properties of the system can be determined.

2.2.2 Theorem II

Theorem II states that a universal functional for the energy $E[n]$ in terms of the density $n(\mathbf{r})$ must exist, valid for any external potential $V(\mathbf{r})$, and the exact ground state energy of the system is the global minimum of this functional. In addition, the density $n(\mathbf{r})$ that minimises this universal functional is the exact ground state density $n_0(\mathbf{r})$.

It can be readily proven from the first theorem by first considering the total

energy as a functional of the charge density, eq. (2.13). Since all the terms in the energy functional are uniquely determined for a given $n(\mathbf{r})$, except for the external interaction, we can construct a functional which encapsulates all internal energies, kinetic and potential of the system. Such a functional is known as the Hohenberg-Kohn (HK) universal functional,

$$F_{\mathrm{HK}}[n] = T[n] + E_{\mathrm{ee}}[n]. \tag{2.14}$$

Hence the final form of the energy functional is given by

$$E_{\mathrm{HK}}[n] = F_{\mathrm{HK}}[n] + \int V(\mathbf{r}) n(\mathbf{r})\, d\mathbf{r}\,. \tag{2.15}$$

Let us now consider a system with external potential $V^{(1)}(\mathbf{r})$ which gives the ground state density $n^{(1)}(\mathbf{r})$. The energy of this system is equal to the HK energy functional as per the above discussion,

$$E^{(1)} = E_{\mathrm{HK}}[n^{(1)}] = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle. \tag{2.16}$$

Now consider another system with density $n^{(2)}(\mathbf{r})$, which must correspond to a different wavefunction $\Psi^{(2)}$. It directly follows that the energy $E^{(2)}$ of this state must be greater than $E^{(1)}$, since

$$E^{(1)} = \left\langle \Psi^{(1)} \middle| \hat{H}^{(1)} \middle| \Psi^{(1)} \right\rangle < \left\langle \Psi^{(2)} \middle| \hat{H}^{(1)} \middle| \Psi^{(2)} \right\rangle = E^{(2)}. \tag{2.17}$$

This shows the energy evaluated for the correct ground state density is indeed lower than the energy evaluated for any other density. Moreover, it implies that minimising $E[n]$ with respect to $n(\mathbf{r})$ would yield the exact ground state density and hence all other ground state observables.

2.3 Kohn-Sham Density Functional Theory

The HK theorems form the basic premise of DFT, where the energy is written as a functional of the density and minimised the energy with respect to the density, to produce the ground state density and energy. The idea of DFT simplifies the original many-body problem greatly. However, with the exact form of the universal functional unknown, this still yields very little applicability to practical calculations. This problem was later ameliorated by Kohn and Sham [22], when they found a remarkable result that the exact ground state density of an interacting system can be calculated using a fictitious non-interacting system. We will now proceed to show

this result and how it can be used to greatly simplify DFT calculations.

Let us begin by considering the universal functional from eq. (2.14), but writing it in terms of a non-interacting system,

$$F[n] = T_s[n] + E_{\mathrm{H}}[n] + E_{\mathrm{xc}}[n]. \tag{2.18}$$

$T_s[n]$ and $E_{\mathrm{H}}[n]$ are the non-interacting kinetic energy and the Hartree energy from the Hartree potential given by eq. (2.6), respectively. $E_{\mathrm{xc}}[n]$ is the exchange-correlation functional that captures the exchange and correlation effects that are missing from the non-interacting electrons, which by definition is equal to

$$E_{\mathrm{xc}}[n] = (T[n] - T_s[n]) + (E_{\mathrm{ee}}[n] - E_{\mathrm{H}}[n]). \tag{2.19}$$

This looks like a small contribution as it only contains the difference in energy, but it has been shown that for homogeneous electron gas, this contribution is of the same order of magnitude as the total kinetic energy [23], thus it cannot be neglected and a good approximation for it is vital for any accurate DFT calculations. Using this form of $F[n]$, $E[n]$ can now be minimised with respect to variations in $n(\mathbf{r})$. This is achieved by solving the Euler-Lagrange equation by imposing $\frac{\delta F[n]}{\delta n} = 0$ along with the orthonormality constraint $\langle \psi_i | \psi_j \rangle = \delta_{ij}$, which results in a one-particle Schrödinger equation

$$\left(-\frac{1}{2}\boldsymbol{\nabla}^2 + V_{\mathrm{eff}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}), \tag{2.20}$$

where the effective potential, $V_{\mathrm{eff}}(\mathbf{r})$, is given by

$$V_{\mathrm{eff}}(\mathbf{r}) = V(\mathbf{r}) + V_{\mathrm{H}}(\mathbf{r}) + V_{\mathrm{xc}}(\mathbf{r}) \tag{2.21}$$

with

$$V_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[n]}{\delta n(\mathbf{r})}, \tag{2.22}$$

and $V_{\mathrm{H}}(\mathbf{r})$ is the Hartree potential from eq. (2.6). Solving the Schrödinger equation would give the density $n(\mathbf{r})$ using the usual definition

$$n(\mathbf{r}) = \sum_i^N f_i |\psi_i(\mathbf{r})|^2, \tag{2.23}$$

where f_i are the occupancies of each state, which is given by Fermi-Dirac distribution.

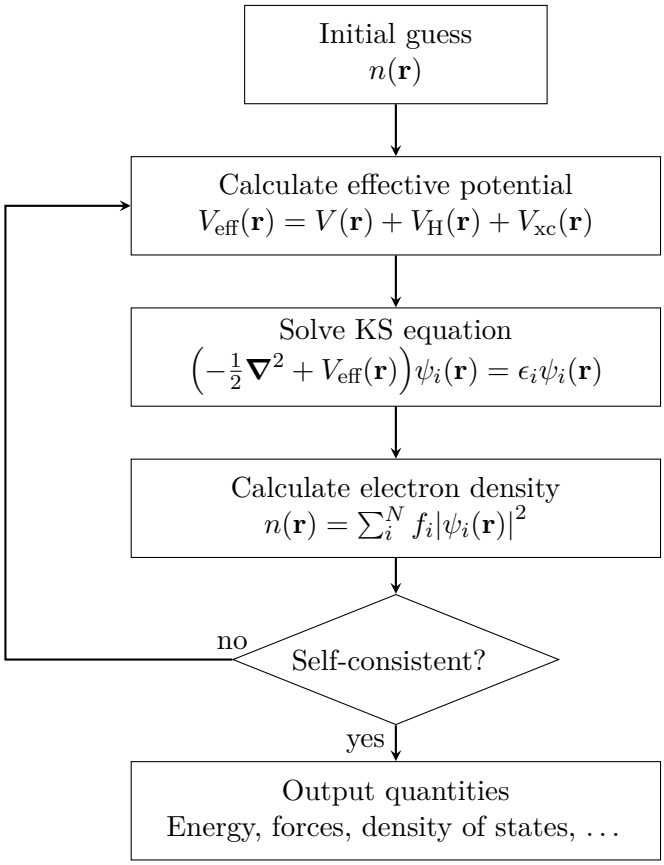


FIGURE 2.1: Schematic representation of self-consistent field calculation of the KS equation.

Equations (2.20) to (2.23) form a set of equations known as the Kohn-Sham (KS) equations. It can be seen that although the derivation of the KS equations required $F[n]$ to be rewritten, it is ultimately unchanged, therefore according to HK theorem II, the density of the non-interacting problem is precisely the same as that of the original interacting problem. In other words, the problem of a system of interacting electrons in an external potential $V(\mathbf{r})$ is equivalent to the problem of a system of non-interacting electrons under the effective potential $V_{\text{eff}}(\mathbf{r})$. From the KS equations, notice how the charge density depends on the effective potential and the effective potential in turn depends on the charge density. This means that the KS equations are a set of closed equations and must be solved self-consistently, as illustrated in fig. 2.1. Needless to say, the exact form of E_{xc} is still unknown, but the problem is now reduced from finding a general universal functional $F[n]$ for an interacting system to just finding a good approximation for E_{xc} for a non-interacting system.

2.4 Exchange-Correlation Functional

Developing a good approximation for the exchange-correlation functional is one of the core problems of DFT. From eq. (2.19), we can see that E_{xc} is numerically small as it is only a small error introduced by the usage of an average Hartree potential and a non-interacting kinetic energy. However, a good approximation for it is vital for any accurate DFT calculations. Within the same seminal paper of Kohn and Sham [22], a very successful approximation, now known as the local density approximation (LDA), was proposed.

For $n(\mathbf{r})$ varying slowly, LDA assumes that the exchange-correlation energy density is the same as in a homogeneous electron gas (HEG). By also separating out the exchange part and correlation part, the LDA functional is written as

$$E_{\text{xc}}^{\text{LDA}}[n] = \int n(\mathbf{r}) \Big(\epsilon_{\text{x}}^{\text{HEG}}(n) + \epsilon_{\text{c}}^{\text{HEG}}(n) \Big) \, d\mathbf{r} \,. \tag{2.24}$$

The exchange part of the functional has the analytical form, following the Thomas-Fermi-Dirac model [24],

$$\epsilon_{\text{x}}^{\text{HEG}}(n) = -\frac{3}{4} \left(\frac{3n(\mathbf{r})}{\pi} \right)^{\frac{1}{3}} . \tag{2.25}$$

The correlation part has been studied analytically by Gell-Mann and Bruecker [25].

Despite the local approximation, LDA is remarkably successful in describing the ground state properties of many real materials. The key to its the success is mostly the fact that the approximation satisfies the sum rules for the exchange-

correlation hole in the charge density [26]. However successful the LDA is, it still has many limitations and net error [27] (See reference [24, 28] for a more in-depth review of the applications of the LDA to real systems and the DFT formulations). An obvious problem with the LDA is that it fails for non-uniform systems where the density can vary rapidly. The generalised gradient approximation (GGA) [29] corrects this by the inclusion of the dependency on the gradient of the density $\boldsymbol{\nabla} n$ and the general form of the functional is

$$E_{\mathrm{xc}}^{\mathrm{GGA}}[n] = \int n(\mathbf{r}) \epsilon_{\mathrm{xc}}^{\mathrm{GGA}}(n, |\boldsymbol{\nabla} n|)\, d\mathbf{r}\,. \tag{2.26}$$

The GGA offers other improvements over LDA in other aspects aside from the rapid varying density problem [30, 31]. Popular functionals that is a family of the GGA includes the PW91 [29], B88 [32] and PBE [33] functionals.

For vertical heterostructures where van der Waals (vdW) forces are present, special care is needed, as GGA and LDA do not include any terms relating to such forces. The vdW interaction is caused by instantaneous dipoles due to correlations in fluctuating polarisations of nearby particles. This effect is long-range and thus vdW forces are non-local in nature. Dion, Rydberg, Schröder, Langreth, and Lundqvist [34] attempted to correct the energy by adding a non-local term in the correlation energy using the LDA as a starting point:

$$E_{\mathrm{c}}[n] = E_{\mathrm{c}}^{\mathrm{LDA}}[n] + E_{\mathrm{c}}^{\mathrm{nl}}[n], \tag{2.27}$$

with the non-local contribution takes the form of a density-density interaction along with a kernel $\phi(\mathbf{r}, \mathbf{r}')$:

$$E_{\mathrm{c}}^{\mathrm{nl}} = \frac{1}{2} \int \int n(\mathbf{r}) n(\mathbf{r}') \phi(\mathbf{r}, \mathbf{r}')\, d\mathbf{r}\, d\mathbf{r}'\,. \tag{2.28}$$

The form of the kernel is generally not known, but can be derived from adiabatic connection-fluctuation theory [35, 36]. This completes the inclusion of vdW forces and the full exchange-correlation energy is simply

$$E_{\mathrm{xc}}[n] = E_{\mathrm{x}}^{\mathrm{GGA}}[n] + E_{\mathrm{c}}^{\mathrm{LDA}}[n] + E_{\mathrm{c}}^{\mathrm{nl}}[n]. \tag{2.29}$$

Unfortunately, this method involves a 6-dimensional integral which is very expensive to compute. However, we can reformulate this problem using a method developed by Román-Peréz and Soler [37] to reduce this to just a 3-dimensional integral. They

rewrite $E_{\mathrm{c}}^{\mathrm{nl}}$ from eq. (2.28) as

$$E_{\mathrm{c}}^{\mathrm{nl}} = \frac{1}{2}\int\int n(\mathbf{r})n(\mathbf{r}')\phi(q,q',\Delta r)\,d\mathbf{r}\,d\mathbf{r}'\,, \tag{2.30}$$

where q and q' are the values at $\mathbf{r}$ and $\mathbf{r}'$ of a universal functional $q_0(n(\mathbf{r}),|\boldsymbol{\nabla} n(\mathbf{r})|)$, and $\Delta r = |\mathbf{r}-\mathbf{r}'|$. The form of the universal functional is proposed in the same paper [37]. We will now consider a fixed number of non-uniform sampling points for the universal functional q_α accompanied by interpolating polynomials $p_\alpha(q)$ to compensate for the finiteness of the grid, where the interacting polynomials are cubic splines with the property $p_\alpha(q_\beta) = \delta_{\alpha\beta}$. With these definitions, the kernel can be written as

$$\phi(q,q',\Delta r) = \sum_{\alpha,\beta} p_\alpha(q)p_\beta(q')\phi_{\alpha\beta}(\Delta r), \tag{2.31}$$

and hence the non-local energy contribution is now

$$E_{\mathrm{c}}^{\mathrm{nl}} = \frac{1}{2}\sum_{\alpha\beta}\int\int n(\mathbf{r})p_\alpha(q)n(\mathbf{r}')p_\beta(q')\phi_{\alpha\beta}(\Delta r)\,d\mathbf{r}\,d\mathbf{r}' \tag{2.32}$$

$$= \frac{1}{2}\sum_{\alpha\beta}\int\int \theta_\alpha(\mathbf{r})\theta_\beta(\mathbf{r}')\phi_{\alpha\beta}(\Delta r)\,d\mathbf{r}\,d\mathbf{r}'\,. \tag{2.33}$$

By performing a Fourier transform and using the Kronecker delta properties, it can be shown that

$$E_{\mathrm{c}}^{\mathrm{nl}} = \frac{1}{2}\sum_{\alpha,\beta}\int \tilde{\theta}_\alpha^*(\mathbf{k})\tilde{\theta}_\beta(\mathbf{k})\tilde{\phi}_{\alpha\beta}(k)\,d\mathbf{k}\,, \tag{2.34}$$

where the '$\sim$' accent denotes the Fourier transform of the function. The original 6-dimensional integral in real space is now reduced to a 3-dimensional integral in reciprocal space, and since the matrix terms introduced by the finite non-uniform grid need only be computed once at the start of the calculation, this formulation improves the computational efficiency greatly compared to the original formulation by Dion et al.

The final problem is now the choice of the functional type, in which Klimeš, Bowler, and Michaelides [38] have explored a variety of exchange functional and the parameters thereof, and have achieved a 'chemical' accuracy within 1kcal/mol for energies for a wide range of systems. They have created special functionals such as the optB88-vDW [38], which will be shown to produce the best results for our systems. The PBE and optB88-vDW functionals are widely used across the results sections of this thesis.

2.5 The Band Gap Problem

Both GGA and LDA possess a fundamental problem where the functional is approximated to be a smooth function, but in reality it is not differentiable in open systems. A proof by Perdew, Parr, Levy, and Balduz shows that E_{xc} must contain a derivative discontinuity [39, 40], because the exact exchange-correlation functional connects the energies for integer particle numbers by straight lines, thus features derivative discontinuities Δ_{xc} at integer particle numbers.

If we consider the fundamental band gap for a system with N electrons is given by the differences of ground state total energies of systems with deviating numbers of electrons [40],

$$E_g = (E_{N+1} - E_N) - (E_N - E_{N-1}). \tag{2.35}$$

Thus being able to calculate the ground-state total energy for these different systems should be enough to calculate the band gap. Within the KS DFT formalism, we can write a similar equation using functional derivatives [40]

$$E_g = \lim_{\eta \to 0^+} \left\{ \left. \frac{\delta E[n]}{\delta n(\mathbf{r})} \right|_{N+\eta} - \left. \frac{\delta E[n]}{\delta n(\mathbf{r})} \right|_{N-\eta} \right\}. \tag{2.36}$$

By employing Janak's theorem [41] ($\frac{\partial E}{\partial n_i} = \epsilon_i$, where ϵ_i is the energy of ith electron state in the KS system) and the derivative discontinuity Δ_{xc} introduced by E_{xc}, we would arrive at

$$E_g = \epsilon_{N+1} - \epsilon_N + \Delta_{\text{xc}}. \tag{2.37}$$

Notice that this result does not assume the form of E_{xc}, so even given the exact exchange-correlation functional, one cannot get the fundamental band gap of the real interacting-electron system without including the derivative discontinuity. This is known as the band gap problem, where DFT always underestimate the band gap. This motivates the development of a different approach that is beyond the KS formalism, called the GW approximation, which is introduced in chapter 4.

2.6 Basis set and Periodic Systems

The set of KS equations (eqs. (2.20) to (2.23)) can now be solved numerically for a given approximation to the exchange-correlation functional by expanding each

one-electron wavefunction as

$$|\psi_i\rangle = \sum_{b=1}^{\infty} |\chi_b\rangle \, c_{bi}. \tag{2.38}$$

$\{\chi_b\}$ are known as basis functions, which form a basis set, and $\{c_{bi}\}$ are the expansion coefficients. The choice of the basis set is free, with one condition that any wavefunction must be able to be written as a linear combination of such basis functions. The expansion in its current form is unusable due to the summation to infinity, therefore an approximation will be needed. A reasonable approximation is simply to reduce the number of basis functions to a finite number, so that the expansion becomes

$$|\psi_i\rangle = \sum_{b=1}^{N_b} |\chi_b\rangle \, c_{bi}, \tag{2.39}$$

where N_b is the number of basis functions. As one can imagine, the larger the basis set used, the higher the accuracy of a DFT calculation at the cost of more computational memory and time. The choice of basis set is therefore twofold: one must choose the type of basis function and the size of the basis set to be used.

Choosing the type of basis set depends on the material in question. This thesis concentrates on crystalline materials, in which the system can be modelled as an infinite repetition of a crystal. This is usually described in terms of a Bravais lattice such that each crystal unit cell are spanned by a set of primitive vectors $\{\mathbf{a}_i\}$ and the infinite repetition can be generated by a discrete translation operation

$$\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{2.40}$$

where n_i are any integers. In such periodic system, the KS effective potential meets the condition

$$v_{\text{eff}}(\mathbf{r}) = v_{\text{eff}}(\mathbf{r} + \mathbf{R}), \tag{2.41}$$

Under such condition, the wavefunction can be written as the Bloch wavefunction, from Bloch's theorem [42],

$$\psi_{i\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{i\mathbf{k}}(\mathbf{r}), \tag{2.42}$$

where $u_{i\mathbf{k}}(\mathbf{r})$ is a periodic function with the same periodicity of the Bravais lattice, so that it possesses the property $u_{i\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{i\mathbf{k}}(\mathbf{r})$. Due to the plane-wave part of the Bloch wavefunction, it is instructive to expand $u_{i\mathbf{k}}(\mathbf{r})$ also in terms of plane-waves.

The Bloch wavefunction therefore becomes

$$\psi_{i\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} c_{i,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \tag{2.43}$$

As stated previously in this section, we must reduce the size of the basis set to a finite number. In this case, we will need to reduce the sum over the reciprocal vectors $\mathbf{G}$ by some spherical cutoff radius, which we will denote as $\mathbf{G}_{\text{cutoff}}$. Note that each plane-wave has a kinetic energy of $\frac{1}{2}|\mathbf{k}+\mathbf{G}|^2$, thus the choice for the $\mathbf{G}_{\text{cutoff}}$ can be defined by a kinetic energy cutoff such that

$$E_{\text{cutoff}} = \frac{1}{2}|\mathbf{G}_{\text{cutoff}}|^2. \tag{2.44}$$

Using a E_{cutoff} to define $\mathbf{G}_{\text{cutoff}}$ is very intuitive because the plane-waves with smaller kinetic energy, electrons that are far away from its nucleus, play a more important role than those with high kinetic energy, electrons that are close to its nucleus.

The plane-wave basis set has advantages such as the lack of the basis set superposition error [43] and due to the independence of the nuclear positions, it allows for simpler calculation of the forces which is described in section 2.8. Such a basis set is used in programs such as QUANTUM ESPRESSO [44], which is used extensively throughout chapter 6.

2.7 Pseudopotentials

Another problem that DFT faces is the complicated effects that core electrons, electrons near its nucleus, experience caused by the divergence of the coulomb potential. One can increase the basis set size to include the higher energy nodes, but this increases the computational cost and can sometimes be impractical. Pseudopotentials essentially remove these core electrons and thus the complicated effects by fixing these electrons as they are so tightly bounded to the nucleus. The potential from the core electrons is replaced by an effective potential that matches the real potential outside of a specified core radius r_c. Similarly, the wavefunction associated with the pseudopotential, the pseudowavefunction, matches the real wavefunction outside of r_c. The pseudopotential is said to be softer with larger r_c.

The choice for the number of core electrons and valence electrons is not strict, but in general follows a simple rule where the outermost shell of an atom is used for the valence electrons and the rest are considered as core electrons. For example, in tungsten one might consider only the $6s^25d^4$ as valence electrons, but there are cases where one needs to include more electronic states as valence: in this case its

semi-core shell $4f^{14}$ may also be included. Including more valence electrons will increase computational cost, but generally improve accuracy; therefore the choice is dependent on the level of accuracy that one is trying to achieve.

There is also another important property that one must consider: the phase shift due to scattering of a plane wave by the core. There is a different phase shift for different angular momentum states, and it is also desirable to reproduce this within the pseudopotential framework. The phase shifts for the valence eigenvalues/eigenfunctions are by definition always matched, however for higher energies this is not the case and other techniques need to be employed to preserve the phase shifts.

The most common type of pseudopotential is the norm-conserving pseudopotential. These pseudopotentials require the norm-conservation condition, where the integrated charge inside r_c for the pseudowavefunction ψ_{PS} must be the same as the real wavefunctions ψ:

$$\int_0^{r_c} |\psi(r)|^2 \, dr = \int_0^{r_c} |\psi_{\mathrm{PS}}(r)|^2 \, dr \, . \tag{2.45}$$

This condition ensures that the energy derivative of the phase shift is the same as for the real wavefunction, thus the scattering phase shift matches the real wavefunction for a wider energy range. Another popular type of pseudopotential is the Vanderbilt ultrasoft pseudopotential [45], where the norm-conservation condition is relaxed in order to reduce the cutoff energy required for the basis set.

Up till now relativistic effects were not discussed, but to encapsulate the effects of spin-orbit coupling (SOC), we must include the effects of relativity into the pseudopotential. Let us start by defining an l-dependent potential for both total angular momentum $j = l + 1/2$ and $j = l - 1/2$ [46, 47]

$$V_l = \frac{l}{2l+1}\Big[(l+1)V_{l+1/2} + lV_{l-1/2}\Big], \tag{2.46}$$

which includes scalar relativistic effects, where the SOC term from the Dirac's equation is ignored. Spin-orbit effects can then be included by adding a short-range non-local term [46]

$$\delta\hat{V}_{\mathrm{SL}}^{\mathrm{so}} = \sum_{lm} |Y_{lm}\rangle \, \delta V_l^{\mathrm{so}}(r) \mathbf{L}\cdot\mathbf{S} \, \langle Y_{lm}| \, , \tag{2.47}$$

where

$$\delta V_l^{\mathrm{so}} = \frac{2}{2l+1}\Big[V_{l+1/2} - V_{l-1/2}\Big]. \tag{2.48}$$

The inclusion of SOC allows the observation of effects such as spin-orbit splitting,

but increases the computational cost. Pseudopotentials generated using the Dirac equation with and without SOC are called fully relativistic pseudopotentials and scalar relativistic pseudopotentials, respectively.

2.8 Forces and The Hellmann-Feynman Theorem

Once the energy is obtained in a DFT calculation one can calculate other properties such as the interatomic forces acting on each atom. Finding the forces of a system is particularly useful as minimising the interatomic forces also minimises the energy, generating the optimal structure from an arbitrary state. However, note that in practice the forces will in general be at a local minimum, because to guarantee a global minimum one is required the search through whole energy space. Such force minimisation is known as geometry optimisation, which allows predicting bond lengths, bond angles and how defects can affect the structure.

The force acting on nucleus I with position R_I can be calculated using

$$\mathbf{F}_I = -\frac{dE}{d\mathbf{R}_I}. \tag{2.49}$$

This derivative can be computed efficiently by employing the Hellmann-Feynman theorem [48, 49]. The theorem states that for a time-independent Schrödinger equation $\hat{H}(\lambda)\psi(\lambda) = E(\lambda)\psi(\lambda)$, the derivative of the energy with respect to the parameter λ is given by

$$\frac{dE(\lambda)}{d\lambda} = \langle\psi|\frac{d\hat{H}}{d\lambda}|\psi\rangle\,. \tag{2.50}$$

The proof of this statement is relatively simple. Let us write the energy derivative in full, giving

$$\frac{dE(\lambda)}{d\lambda} = \left\langle\frac{d\psi}{d\lambda}\middle|\hat{H}\middle|\psi\right\rangle + \left\langle\psi\middle|\hat{H}\middle|\frac{d\psi}{d\lambda}\right\rangle + \langle\psi|\frac{d\hat{H}}{d\lambda}|\psi\rangle \tag{2.51}$$

$$= E(\lambda)\left[\left\langle\frac{d\psi}{d\lambda}\middle|\psi\right\rangle + \left\langle\psi\middle|\frac{d\psi}{d\lambda}\right\rangle\right] + \langle\psi|\frac{d\hat{H}}{d\lambda}|\psi\rangle\,. \tag{2.52}$$

Since the wavefunction must obey the condition $\langle\psi|\psi\rangle = 1$, we can write

$$\frac{d}{d\lambda}\langle\psi|\psi\rangle = \left\langle\frac{d\psi}{d\lambda}\middle|\psi\right\rangle + \left\langle\psi\middle|\frac{d\psi}{d\lambda}\right\rangle = 0. \tag{2.53}$$

Hence,

$$\frac{dE(\lambda)}{d\lambda} = \langle\psi|\frac{d\hat{H}}{d\lambda}|\psi\rangle\,. \tag{2.54}$$

In other words, the energy derivative with respect to an external parameter can be calculated as the expectation value of the derivative of the Hamiltonian operator. The force acting on nucleus I can therefore simply be calculated using

$$\mathbf{F}_I = -\langle\psi|\frac{d\hat{H}}{d\mathbf{R}_I}|\psi\rangle\,. \tag{2.55}$$

A problem to note is that the Hellmann-Feynman theorem requires that the wavefunction to be in its exact ground state, which is not possible to obtain with a variational method such as DFT. At each iteration, the varying wavefunction, which we will call ψ^{trial}, will not be the eigenfunction of the Hamiltonian, that is $\hat{H}\psi^{\text{trial}} \neq E\psi^{\text{trial}}$. Therefore, the Hellmann-Feynman theorem cannot be applied directly. We must consider the full energy derivative for a wavefunction which is expanded in terms of a basis set $|\psi\rangle = \sum_b |\chi_b\rangle\, c_b$. Using chain rule, the energy derivative gives

$$\mathbf{F}_I = -\frac{dE}{d\mathbf{R}_I} = -\frac{\partial E}{\partial \mathbf{R}_I} - \sum_b \frac{\partial E}{\partial c_b}\frac{\partial c_b}{\partial \mathbf{R}_I} - \sum_b \int \frac{\partial E}{\partial \chi_b(\mathbf{r})}\frac{\partial \chi_b(\mathbf{r})}{\partial \mathbf{R}_I}\, d\mathbf{r}\,. \tag{2.56}$$

The first term is just the Hellmann-Feynman force. The second and third terms are known as the Pulay force [50]. Just looking at the Pulay force, when full self-consistent convergence to the exact ground state is achieved, the second term will vanish because $\frac{\partial E}{\partial c_b} = 0$. The third term can only vanish when $\frac{\partial \chi_b(\mathbf{r})}{\partial \mathbf{R}_I}$ vanishes, because $\frac{\partial E}{\partial \chi_b(\mathbf{r})}$ is non-zero unless the basis set is complete, i.e., infinitely large. Since a plane-wave basis set does not depend on the nuclear positions, $\frac{\partial \chi_b(\mathbf{r})}{\partial \mathbf{R}_I} = 0$, thus the Pulay force does not contribute to the total force of an atom and the Hellmann-Feynman theorem can be applied directly.

Chapter 3

Linear-Scaling Density Functional Theory

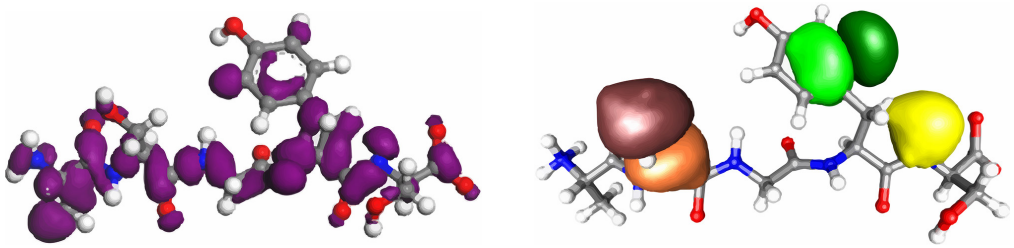
3.1 Limitations of Traditional DFT

Introducing the formalism of DFT has indeed simplified the original many-body problem, where calculation of only a few atoms is very cumbersome, and has thus made enormous impact on science [51]. However, traditional KS DFT still suffers a major problem, namely that simulation of large systems beyond a few hundred atoms is difficult even for the most powerful supercomputers. The origin of this problem lies in the $O(N^3)$ scaling of DFT, where as the number of atoms increases, the computational cost increases cubically. To understand this $O(N^3)$ asymptotic behaviour, let us go back to the KS equation, eq. (2.20). As can be seen from this equation, the orthonormality constraint of the KS orbitals must be imposed directly,

$$\int \psi_m^*(\mathbf{r})\psi_n(\mathbf{r})\, d\mathbf{r} = \delta_{nm}. \tag{3.1}$$

As the number N of KS orbitals rises, the number of orbital pairs grows as N^2. Furthermore, as illustrated in fig. 3.1a, each KS orbital in general extends over the entire system, thus the overlap integral in eq. (3.1) scales with the system size as $O(N)$. Combining all the aforementioned dependencies, a traditional KS DFT calculation requires a computational effort which scales as $O(N^3)$. Therefore, in order to study larger and more complex systems, this time complexity will need to be reduced.

Recent years have shown an increase of interest in the development of linear-scaling or $O(N)$ DFT [54, 55], which offers the power to perform simulations of



(a) (b)

FIGURE 3.1: Orbitals for a peptide from two different DFT calculations. (a) A traditional DFT calculation with CASTEP [52] showing a delocalised KS orbital. (b) A ONETEP calculation showing three localised orbitals (non-orthogonal generalised Wannier functions). Courtesy of Skylaris et al. [53]

entire biological molecules and nanostructures containing thousands of atoms. There have been several different linear-scaling schemes proposed [56–60], but this project focuses on a linear-scaling method that has been implemented in the ONETEP code [53, 61].

3.2 Density Matrix Formulation

We shall now consider the density matrix and transform it to a form appropriate to a linear-scaling approach. Within KS-DFT a spinless reduced single particle density matrix is defined as

$$\hat{\rho} = \sum_n f_n \left|\psi_n\right\rangle\!\left\langle\psi_n\right|, \tag{3.2}$$

where f_n is the occupation number, which is either 0 or 1 for an insulating system. This density matrix possesses the property of idempotency,

$$\begin{aligned}\rho^2(\mathbf{r},\mathbf{r}') &= \sum_{ij} f_i f_j \left\langle\mathbf{r}|\psi_i\right\rangle \left\langle\psi_i|\psi_j\right\rangle \left\langle\psi_j|\mathbf{r}'\right\rangle = \sum_{ij} f_i f_j \left\langle\mathbf{r}|\psi_i\right\rangle \delta_{ij} \left\langle\psi_j|\mathbf{r}'\right\rangle \\ &= \sum_k f_k \left\langle\mathbf{r}|\psi_k\right\rangle \left\langle\psi_k|\mathbf{r}'\right\rangle = \rho(\mathbf{r},\mathbf{r}'),\end{aligned} \tag{3.3}$$

which requires the orthonormality of the orbitals $\{\psi_n(\mathbf{r})\}$ as per eq. (3.1). Because of this property of idempotency, the density matrix is thus a projection operator, in this case, an operator that projects a state onto the space of the occupied states $\hat{\rho}$.

The diagonal elements of $\rho(\mathbf{r},\mathbf{r}')$ give the charge density $n(\mathbf{r})$, that is to say

$$n(\mathbf{r}) = 2\rho(\mathbf{r},\mathbf{r}). \tag{3.4}$$

The factor of 2 accounts for the spin degeneracy in our closed-shell system. This

also means that the total number of electrons must be given by trace of ρ,

$$2\,\mathrm{tr}[\rho] = N. \tag{3.5}$$

An important property of the density matrix is that the total energy of the system may be written in terms of the trace of the product of the Hamiltonian and density,

$$E_{\mathrm{tot}} = \mathrm{tr}\left[\hat{\rho}\hat{H}\right]. \tag{3.6}$$

As it is now, the information encompassed in the density matrix still scales as N^2 because the orbitals are delocalised and the number of states scales linearly, thus the original scaling problem is not solved.

To proceed further in the development of linear-scaling, one must make some approximations, namely exploiting the nearsightedness principle [62, 63]. It has been proven analytically [64] that in insulating or semiconducting systems, the density matrix decays exponentially with respect to the displacement between the atoms, so that

$$\rho(\mathbf{r}, \mathbf{r}') \propto e^{-\gamma|\mathbf{r}-\mathbf{r}'|} \to 0 \quad \text{as} \quad \left|\mathbf{r}-\mathbf{r}'\right| \to \infty. \tag{3.7}$$

This means that for any given position $\mathbf{r}$ the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ differs significantly from zero only for points $\mathbf{r}'$ within a finite volume around $\mathbf{r}$. Note that this volume is independent of system-size, as the decay rate γ depends on the energy gap between the highest occupied and lowest unoccupied states. Hence, the nearsightedness principle indicates that the density matrix can be expressed as a sparse matrix. When written in a localised basis, this implies that the total amount of important information scales linearly with system size. Furthermore, the use of localised orbitals will not result in lost accuracy.

Within ONETEP, so-called non-orthogonal generalised Wannier functions (NGWFs) are used as the localised orbital [53, 65] (illustrated in fig. 3.1b) and the density matrix is written in a separable form,

$$\rho(\mathbf{r}, \mathbf{r}') = \sum_{\alpha,\beta} \phi^*_\alpha(\mathbf{r}) K^{\alpha\beta} \phi_\beta(\mathbf{r}'), \tag{3.8}$$

where $\phi_\alpha(\mathbf{r})$ is the NGWF and $K^{\alpha\beta}$ is known as the density kernel [66]. To understand the density kernel, let us relate the NGWFs $\phi_\alpha(\mathbf{r})$ back to the more familiar KS orbitals $\psi_i(\mathbf{r})$:

$$\psi_i(\mathbf{r}) = \sum_\alpha M_i{}^\alpha \phi_\alpha(\mathbf{r}), \tag{3.9}$$

where ${M_i}^\alpha$ is the molecular orbital coefficient of orbital i corresponding to NGWF ϕ_α. With the NGWFs in this form, they can be truncated beyond a certain cutoff radius R_ϕ. Using this relationship along with eq. (3.2) and eq. (3.8) will give the definition of the density kernel,

$$K^{\alpha\beta} = \sum_n M_n^{\dagger\alpha} f_n {M_n}^\beta. \tag{3.10}$$

With the new definition of the density matrix, the idempotency condition proposed in eq. (3.3) will also need to be reformulated. $\rho^2(\mathbf{r}, \mathbf{r}')$ now reads

$$\begin{aligned} \rho^2(\mathbf{r}, \mathbf{r}') &= \sum_{\alpha,\delta} \langle \mathbf{r}|\phi_\alpha\rangle \, K^{\alpha\delta} \, \langle \phi_\delta| \left[\sum_{\gamma,\beta} |\phi_\gamma\rangle \, K^{\gamma\beta} \, \langle \phi_\beta|\mathbf{r}'\rangle \right] \\ &= \sum_{\alpha,\beta} \langle \mathbf{r}|\phi_\alpha\rangle \left[\sum_{\delta,\gamma} K^{\alpha\delta} S_{\delta\gamma} K^{\gamma\beta} \right] \langle \phi_\beta|\mathbf{r}'\rangle \,, \end{aligned} \tag{3.11}$$

where $S_{\delta\gamma} = \langle \phi_\delta|\phi_\gamma\rangle$ is the overlap integral between NGWFs ϕ_δ and ϕ_γ. It is obvious that for the idempotency condition to be satisfied, the term inside the square brackets must equal to the corresponding term of the density kernel, i.e.,

$$\mathbf{KSK} = \mathbf{K}. \tag{3.12}$$

The normalisation condition, eq. (3.5) can also be redefined as,

$$\begin{aligned} 2\,\mathrm{tr}[\rho] &= 2 \sum_{\alpha,\delta,\gamma} \langle \phi_\alpha|\phi_\delta\rangle \, K^{\delta\gamma} \, \langle \phi_\gamma|\phi^\alpha\rangle \\ &= 2 \sum_{\alpha,\gamma,\delta} S_{\alpha\delta} K^{\delta\gamma} {\delta_\gamma}^\alpha = 2\,\mathrm{tr}\left[\mathbf{KS}\right] = N_e, \end{aligned} \tag{3.13}$$

where

$$\phi^\alpha = \sum_{\alpha,\beta} \left(S^{-1}\right)^{\alpha\beta} \phi_\beta \tag{3.14}$$

is defined as a dual to the NGWFs, thus possesses the property that $\langle \phi_\beta|\phi^\alpha\rangle = {\delta_\beta}^\alpha$. With the above redefinitions, the sparsity of the density kernel can then be predetermined. Since the density kernel is required to be a sparse matrix, it then needs to be truncated by postulating that $K^{\alpha\beta}$ can be set to 0 if the corresponding NGWFs centres are separated by more than some cutoff r_K.

The localisation of the NGWFs along with the density kernel truncation will result in a density matrix where the information truly scales linearly with the size of the system. Note that this only applies in the case of insulators or semiconductors.

3.3 Density Kernel Optimisation

The idempotency constraint (eq. (3.12)) presented in the last subsection needs to be imposed. There are two separate methods used within ONETEP for this: the penalty-functional method [66, 67] and the method of Li, Nunes and Vanderbilt (LNV) [68], based on the McWeeny purification transformation [66]. The first method imposes the idempotency by first minimising the positive definite functional $P[\rho]$ using a steepest descent method, where

$$P[\rho] = \int\int \Big[\rho^2(\mathbf{r},\mathbf{r}') - \rho(\mathbf{r},\mathbf{r}')\Big]\Big[\rho^2(\mathbf{r}',\mathbf{r}) - \rho(\mathbf{r}',\mathbf{r})\Big]\, d\mathbf{r}\, d\mathbf{r}' \,. \tag{3.15}$$

For a nearly idempotent density matrix, this is proportional to the mean square deviation of the occupation numbers, i.e.,

$$P[\rho] = \mathrm{tr}\left[(\rho^2 - \rho)^2\right] = \sum_n (f_n^2 - f_n)^2. \tag{3.16}$$

Then a certain proportion of this functional is added to the energy functional to enforce idempotency. This method, however, has the limitation that the functional can have several local minima.

The second method, the default method in ONETEP, by expressing the density-kernel in terms of an auxiliary matrix **L**:

$$\mathbf{K} = 3\mathbf{LSL} - 2\mathbf{LSLSL}, \tag{3.17}$$

where **L** is defined as

$$\mathbf{L}[\{\phi_\alpha\}] = \min_{\{K^{\alpha\beta}\}} E[\{K^{\alpha\beta}\};\{\phi_\alpha\}]. \tag{3.18}$$

Hence, the functional for E is minimised to find $K^{\alpha\beta}$ for a fixed $\{\phi_\alpha\}$. The energy functional for this method only has a single minimum, as opposed to the former method. Optimising the kernel this way is stable for when the occupation numbers or the eigenvalues of **L**, are in the range of $\left[\frac{1-\sqrt{5}}{2}, \frac{1+\sqrt{5}}{2}\right]$.

3.4 Non-Orthogonal Generalised Wannier Functions

At the heart of ONETEP is the choice of the orbitals, in which a set of NGWFs is used. They are a small number of iteratively optimised local orbitals, as opposed to a large fixed set. This allows excellent transferability since it is no longer necessary to

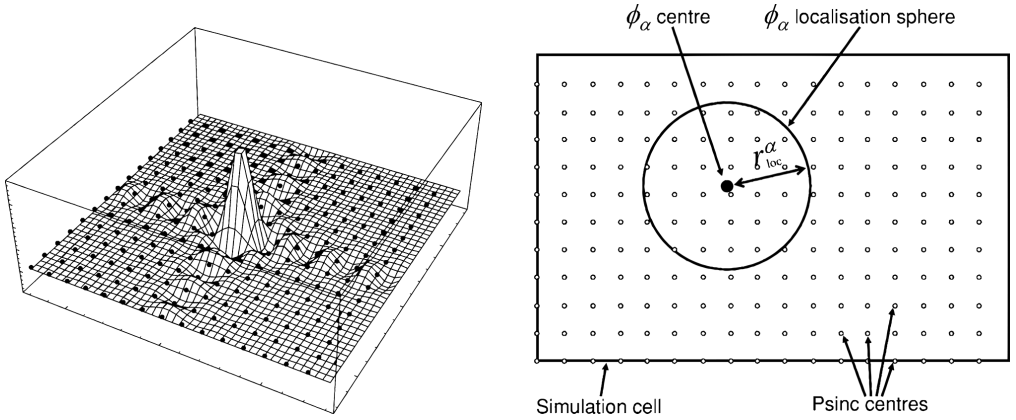


FIGURE 3.2: (a) A psinc basis function centered at a grid point in the simulation cell, where it crosses each grid point (black dots) with value zero, except for the grid point that it is centred on. (b) NGWF expansion in terms of the psinc basis with localisation radius imposed. Courtesy of Skylaris et al [53].

increase the size of the orbital set to reach convergence. Although extra computation is required to optimise the NGWFs, this computational effort is compensated by the minimal orbital set size, while achieving high accuracy.

We will begin by describing the basis set used to expand the NGWFs. Let us consider a simulation cell containing a total number of grid points G, $\mathbf{r}_i$ is a real-space grid point and $\mathbf{k}_p$ is a reciprocal-space grid point. Centered on each of the grid points is a basis function given by

$$D_i(\mathbf{r}) = \frac{1}{G}\sum_p e^{i\mathbf{k}_p\cdot(\mathbf{r}-\mathbf{r}_i)}. \tag{3.19}$$

This is known as a periodic cardinal sine (psinc) function [69] or a periodic bandwidth-limited delta function, as illustrated in fig. 3.2a. The maximum value of $\mathbf{k}_p$ is controlled by the usual user-specified energy cutoff E_{cut}, and is given by

$$\max(\mathbf{k}_p) = \sqrt{2E_{\text{cut}}}. \tag{3.20}$$

Hence the grid spacing can be controlled through this energy cutoff parameter. The psinc functions have the property that they are zero at any other grid point aside from the one which they are centered on, but they do have non-zero values and oscillate between grid points.

Notice that fig. 3.2a lacks cylindrical symmetry. This becomes more apparent when looking at the traditional 2D sinc function, which is written as $\frac{\sin(x)\sin(y)}{xy}$.

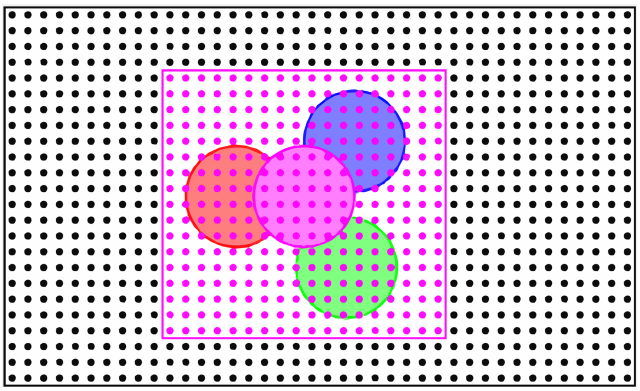


FIGURE 3.3: Illustration of the FFT box technique. The circles are the different NGWFs, the outer box is the simulation cell and the inner box is the region in which the FFTs are performed.

Because the sinc function can be written in terms of a complex integral:

$$\mathrm{sinc}(\mathbf{r}) = \frac{1}{2}\int_{-1}^{1} e^{i\mathbf{t}\cdot\mathbf{r}}\, d\mathbf{t}\,, \tag{3.21}$$

the psinc function differs only in that the continuous integral is replaced by a finite sum over the reciprocal lattice vectors of the simulation cell, as in eq. (3.19). As a result, they are also cell periodic, i.e., $D_i(\mathbf{r}+\mathbf{R}) = D_i(\mathbf{r})$. Since this basis spans the same Hilbert space as the basis of plane-waves via Fourier transform, the kinetic energy can be calculated accurately and efficiently using fast Fourier transforms (FFTs).

Consequently, the NGWFs ϕ_α are expanded in terms of the basis $D(\mathbf{r})$,

$$\phi_\alpha(\mathbf{r}) = \sum_i c_{i,\alpha} D_i(\mathbf{r}) \quad \forall\, i \in R_\alpha, \tag{3.22}$$

where $C_{i,\alpha}$ is the expansion coefficient of ϕ_α at grid point i and R_α is the localisation radius of ϕ_α. ϕ_α outside of R_α is zero as seen in fig. 3.2b. As a result of the NGWFs being written in this form with some localisation radius that is set *a priori*, the FFTs required to calculate the kinetic energy need not to be performed over the whole simulation cell. Instead, FFTs can be performed in boxes with the so-called FFT box technique [70]. A smaller FFT box is placed such that it is centred on each NGWFs and its size is chosen to be sufficiently large, six times the size of the NGWFs radius, to encompass all the NGWFs of the overlapping neighours, see fig. 3.3. FFTs performed this way will depend only on the radius of the NGWFs and not on the system size, thus linearly scaling is maintained.

3.5 Projector Augmented Wave

The projector augmented wave (PAW) method was introduced by Blöchl in 1994 [71] to resolve the difficulty for electronic structure methods to describe the bonding region without the loss of accuracy or speed, while considering the highly oscillating valence wave functions near the atom center. It is a method which generalises the pseudopotential concept and augmented wave methods, and allows for smoother wavefunctions which in turn allows the calculation of all-electron (AE) quantities.

3.5.1 Wavefunctions

As in other augmented wave methods, the PAW method divides the wave function into parts: a partial wave expansion within an atom-centred sphere, and plane waves outside the sphere. The region inside this sphere is known as the augmentation region and this sphere shall be known as the PAW sphere. In order to solve the problem of the strongly oscillating wavefunctions near the atom core, one needs to consider a linear transformation which maps the AE eigenstates $|\psi_n\rangle$ onto some computationally convenient pseudo (PS) wavefunctions $\left|\tilde{\psi}_n\right\rangle$:

$$|\psi_n\rangle = \hat{\mathcal{T}}\left|\tilde{\psi}_n\right\rangle, \tag{3.23}$$

where $\hat{\mathcal{T}}$ is the linear transformation operator. Note that we have defined all the PS quantities to have '$\sim$' as accent. The linear transformation is defined to only act on wavefunctions inside the augmentation region R, so that

$$\hat{\mathcal{T}} = 1 + \sum_R \hat{\mathcal{T}}_R. \tag{3.24}$$

This implies that AE and PS wavefunctions are identical outside R.

If the AE and PS wavefunctions are expanded in a set of partial waves $|\varphi_i\rangle$ and PS partial waves $|\tilde{\varphi}_i\rangle$, such that

$$|\psi\rangle = \sum_i c_i\,|\varphi_i\rangle \quad \text{and} \quad \left|\tilde{\psi}\right\rangle = \sum_i c_i\,|\tilde{\varphi}_i\rangle\,, \tag{3.25}$$

then the full wavefunction in all space can be written as

$$|\psi\rangle = \left|\tilde{\psi}\right\rangle + \sum_i c_i(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle). \tag{3.26}$$

Since $\hat{\mathcal{T}}$ is required to be linear, the expansion coefficients of the partial waves must

be given by a projection in each sphere

$$c_i = \left\langle \tilde{p}^i \middle| \tilde{\psi} \right\rangle, \tag{3.27}$$

for some set of projectors $\tilde{p}$. Combining the above relations, one arrives at

$$|\psi_n\rangle = \hat{\mathcal{T}} \left| \tilde{\psi}_n \right\rangle = \left(1 + \sum_i \left(|\varphi_i\rangle - |\tilde{\varphi}_i\rangle \right) \left\langle \tilde{p}^i \right| \right) \left| \tilde{\psi}_n \right\rangle, \tag{3.28}$$

giving the definition of the linear transformation $\hat{\mathcal{T}}$. The projectors $|\tilde{p}^i\rangle$ are dual to the partial waves $|\tilde{\varphi}_i\rangle$, i.e.,

$$\left\langle \tilde{p}^i \middle| \tilde{\varphi}_j \right\rangle = \delta_{ij}. \tag{3.29}$$

In addition, the PS partial waves should ideally therefore form a complete set such that

$$\sum_i \left| \tilde{\varphi}_i \right\rangle \left\langle \tilde{p}^i \right| = 1. \tag{3.30}$$

In practice, we want the partial waves and the projectors in a form where the radial and angular dependence is separable. Therefore, expanded into a radial part and a spherical part according to their corresponding quantum numbers,

$$\xi_i(\mathbf{r}) = \frac{\xi_{n_i,l_i}(r)}{r} S_{l_i,m_i}(\hat{\mathbf{r}}), \tag{3.31}$$

where ξ_i is a partial wave or projector function.

3.5.2 Densities

Within ONETEP, the eigenstates are not directly accessible, therefore the transformation $\hat{\mathcal{T}}$ must be applied to the AE density matrix. Let us begin with the AE density matrix,

$$\begin{aligned}
\rho &= \sum_n \langle \psi_n | f_n | \psi_n \rangle \\
&= \sum_n \left(\left\langle \tilde{\psi}_n \right| + \sum_i \left(\langle \varphi_i | - \langle \tilde{\varphi}_i | \right) \left\langle \tilde{p}^i \middle| \tilde{\psi}_n \right\rangle \right) f_n \left(\left| \tilde{\psi}_n \right\rangle + \sum_j \left(|\varphi_j\rangle - |\tilde{\varphi}_j\rangle \right) \left\langle \tilde{\psi}_n \middle| \tilde{p}^j \right\rangle \right) \\
&= \sum_n \left\langle \tilde{\psi}_n \middle| f_n \middle| \tilde{\psi}_n \right\rangle + \sum_{ij} \left(\langle \varphi_i | - \langle \tilde{\varphi}_i | \right) \left(|\varphi_j\rangle - |\tilde{\varphi}_j\rangle \right) \sum_n \left\langle \tilde{p}^i \middle| \tilde{\psi}_n \right\rangle f_n \left\langle \tilde{\psi}_n \middle| \tilde{p}^j \right\rangle \\
&\quad + \sum_i \left(\langle \varphi_i | - \langle \tilde{\varphi}_i | \right) \left| \tilde{p}^i \right\rangle \sum_n \left\langle \tilde{\psi}_n \middle| f_n \middle| \tilde{\psi}_n \right\rangle + \sum_n \langle f_n | \tilde{\psi}_n | f_n \rangle \sum_j \left| \tilde{p}^j \right\rangle \left(|\varphi_j\rangle - |\tilde{\varphi}_j\rangle \right).
\end{aligned} \tag{3.32}$$

As can be seen here the PAW transformation in eq. (3.28) has been employed. The above expression can be simplified by the use of the PS density matrix $\tilde{\rho} = \sum_n \left\langle \tilde{\psi}_n \middle| f_n \middle| \tilde{\psi}_n \right\rangle$, the assumption that the PS partial waves form a complete set within the PAW spheres, as in eq. (3.30), and also the assumption that any non-local operators will not act between different PAW spheres. This gives

$$\rho = \tilde{\rho} + \sum_{ij} \left(\langle \varphi_i | \varphi_j \rangle - \langle \tilde{\varphi}_i | \tilde{\varphi}_j \rangle \right) \left| \tilde{p}^i \right\rangle \tilde{\rho} \left\langle \tilde{p}^j \right| . \tag{3.33}$$

Hence, we can write the total AE valence density in terms of the sum of the density associated with the PS wavefunctions and a correction term $\rho^1(\mathbf{r}) - \tilde{\rho}^1(\mathbf{r})$ localised in the spheres around each atom:

$$\begin{aligned} \rho(\mathbf{r}) &= \tilde{\rho}(\mathbf{r}) + \sum_{ij} \rho^{ij} \langle \varphi_i | \mathbf{r} \rangle \langle \mathbf{r} | \varphi_i \rangle - \sum_{ij} \rho^{ij} \langle \tilde{\varphi}_i | \mathbf{r} \rangle \langle \mathbf{r} | \tilde{\varphi}_i \rangle \\ &= \tilde{\rho}(\mathbf{r}) + \rho^1(\mathbf{r}) - \tilde{\rho}^1(\mathbf{r}). \end{aligned} \tag{3.34}$$

The ρ^{ij} matrix is the projection of the density matrix in the PAW spheres:

$$\rho^{ij} = \left| \tilde{p}^i \right\rangle \tilde{\rho} \left\langle \tilde{p}^j \right| . \tag{3.35}$$

Since the PS density $\tilde{\rho}(\mathbf{r})$ is smooth, it can be treated on a grid just like in a norm-conserving approach. $\rho^1(\mathbf{r})$ and $\tilde{\rho}^1(\mathbf{r})$ are changing rapidly, and therefore required to be dealt with on radial grids around each atom. As can be seen here, the total AE density is recovered by applying a general PS wavefunction scheme. ONETEP follows the PAW implementation in ABINIT [72] and we refer the reader to the detailed ONETEP implementation written by Nicholas Hine [73] for any other details related to PAW in linear-scaling DFT.

Chapter 4

The GW Approximation

4.1 Motivations for the GW Approximation

As briefly stated in section 2.5, DFT encounters a fundamental problem when predicting the band gap of a material. The LDA and GGA fail to predict transition metal oxides to be an insulator. A early attempt to mediate this problem was the LDA+U method [74, 75], where an orbital-dependent potential U that acts only on the d or f states is introduced on top of LDA. This introduction of this extra potential splits partially-filled d or f bands, forming the upper and lower Hubbard bands. This was proven to work reasonably well for Mott-Hubbard insulators and rare earth metals, in which the 3d and 4f bands are partially filled [76]. LDA+U, however, incorrectly predicts the band structure and does not work well for metals that contain partially filled 3d shells such as transition metals. Another method to improve the energies is to calculate the exact exchange energy and use the LDA for the correlation energy [77, 78]. This is known as the optimized effective potential method. It has been shown to yield promising results for semiconductors and insulators regarding to the band gaps [79], but since it was intended to only improve upon the energy functional, it does not address the description of quasiparticle (QP) energies, which DFT also lacks.

QP energies are the energies of particle-like excitations of an interacting many-electron system, in contrast to the KS eigenvalues, which have no clear physical meaning as the eigenvalues are of a non-interacting system. QP energies are therefore more useful for excited-state properties, for example, the electronic excitation spectra of materials such as those measured in optical and photoemission experiments. The GW approximation (GWA) addresses the problem with the band gap and QP energies by the introduction of a non-local and energy-dependent self-energy

operator that replaces the exchange-correlation energy functional from DFT Hedin [80]. The GWA is also equipped with a screened Coulomb potential to include the effect of screening which is ideal for electron gas and semiconductors with a large dielectric constant (e.g., silicon with a dielectric constant of ~ 12 [81]), where screening is very important. The GWA has been applied to wide classes of materials and successfully achieved good agreements with experimental band gaps [82]. In this chapter, many formulations and ideas have their origin in the paper by Aryasetiawan and Gunnarsson [83].

4.2 The Green's Function and the Self-Energy

In order to calculate the QP excitation energies, we must consider adding an electron to a higher energy unoccupied state. The excited electron or quasiparticle have a finite lifetime and a different effective mass, but otherwise maintains the same fermionic properties as the bare electron. QP excitation is not to be confused with optical excitation. In QP excitation, we are only adding or subtracting an electron to the system, whereas in optical excitation, an electron is excited from the valence band to the conduction band, which introduces two QPs into the system, the electron and hole, which interact and can form an exciton.

During an excitation, a hole is created which later annihilates with an electron during photoemission. Under such description, it is intuitive to use the annihilation and creation second-quantised operators in the Heisenberg representation. Along with the Green's function (GF) formalism, we can describe the probability of finding a particle at a position r at time t, after adding a particle at position r' at time t' with

$$G(\mathbf{r},t;\mathbf{r}',t') = -i\left\langle \Psi_0^N \middle| T[\hat{\psi}(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t')] \middle| \Psi_0^N \right\rangle . \tag{4.1}$$

$\left|\Psi_0^N\right\rangle$ is the exact N-electron ground state, the $\hat{\psi}(\mathbf{r},t)$ and $\hat{\psi}^\dagger(\mathbf{r},t)$ are the annihilation and creation operators, and T indicates that the operators are time-ordered such that

$$G(\mathbf{r},t;\mathbf{r}',t') = \begin{cases} -i\left\langle \Psi_0^N \middle| \hat{\psi}(\mathbf{r},t)\hat{\psi}^\dagger(\mathbf{r}',t') \middle| \Psi_0^N \right\rangle & \text{for } t > t' \\ i\left\langle \Psi_0^N \middle| \hat{\psi}^\dagger(\mathbf{r}',t')\hat{\psi}(\mathbf{r},t) \middle| \Psi_0^N \right\rangle & \text{for } t < t'. \end{cases} \tag{4.2}$$

To get a more physical description of the GF, we can imagine in a photoemission experiment, one would fire a photon with an energy ω to knock out electrons and measure the energy of the photoemitted electrons to obtain an excitation spectrum. Therefore, it is useful express the GF as a function of frequency, and hence

energy, ω. For that we can perform a Fourier transform with respect to $t - t'$, assuming time-translational invariance, to arrive at the spectral representation of the GF

$$G(\mathbf{r}, \mathbf{r}', \omega) = \sum_n \frac{\left\langle \Psi_0^N \middle| \hat{\psi}(\mathbf{r}) \middle| \Psi_n^{N+1} \right\rangle \left\langle \Psi_n^{N+1} \middle| \hat{\psi}^\dagger(\mathbf{r}') \middle| \Psi_0^N \right\rangle}{\omega - (E_n^{N+1} - E_0^N) + i\delta} + \frac{\left\langle \Psi_0^N \middle| \hat{\psi}^\dagger(\mathbf{r}) \middle| \Psi_n^{N-1} \right\rangle \left\langle \Psi_n^{N-1} \middle| \hat{\psi}(\mathbf{r}') \middle| \Psi_0^N \right\rangle}{\omega - (E_0^N - E_n^{N+1}) + i\delta}, \tag{4.3}$$

where δ is an infinitesimal ($\delta = 0^+$) and $\left|\Psi_n^{N\pm1}\right\rangle$ are the wavefunctions of the many-body Hamiltonian with $N \pm 1$ electrons. The $E_n^{N+1} - E_0^N$ and $E_0^N - E_n^{N+1}$ terms represent the energy of adding an electron to, or removing an electron from, an N-electron system, where $E_n^{N\pm1}$ is the energy of the $N \pm 1$ electrons system of the i-th excited state. This energy is the QP energy, which we will define as

$$\epsilon_n = \begin{cases} E_n^{N+1} - E_0^N & \epsilon_n \geq \epsilon_F \\ E_0^N - E_n^{N-1} & \epsilon_n < \epsilon_F, \end{cases} \tag{4.4}$$

where ϵ_F is the Fermi energy.

The GF from eq. (4.3) is not particularly useful, as the eigenstates of the interacting many-body Hamiltonian is not known. We can, however, write an equivalent GF associated for a non-interacting system and include the electron-electron interactions using perturbation theory. By using the mean-field (MF) wavefunctions and energies, we can write the non-interacting GF as

$$G_0(\mathbf{r}, \mathbf{r}', \omega) = \sum_n \frac{\phi_n^{\mathrm{MF}}(\mathbf{r}) \phi_n^{\mathrm{MF}*}(\mathbf{r}')}{\omega - \epsilon_n^{\mathrm{MF}} \pm i\delta}, \tag{4.5}$$

where $\phi_n^{\mathrm{MF}}(\mathbf{r})$ and ϵ_n^{MF} are the MF wavefunctions and energies, and δ is an infinitesimal that is positive or negative based on whether the state is above or below the Fermi energy. We can now apply a perturbation to this non-interacting GF to obtain a Dyson equation for the single-particle GF

$$G(\mathbf{r}, \mathbf{r}', \omega) = G_0(\mathbf{r}, \mathbf{r}', \omega) + \int G_0(\mathbf{r}, \mathbf{r}_1, \omega) \Sigma(\mathbf{r}_1, \mathbf{r}_2, \omega) G(\mathbf{r}_2, \mathbf{r}', \omega)\, d\mathbf{r}_1\, d\mathbf{r}_2\,. \tag{4.6}$$

Σ is a non-local energy-dependent operator known as the self-energy operator, which contains all of the correlation effects of a many-body system that is missing from a non-interacting system. The self-energy operator is analogous to the exchange-correlation functional in DFT.

The simplest way to approximate Σ is to expand it as a power series, also

called the Dyson series, of the bare Coulomb potential, in which truncating it after the first term would yield the Hartree-Fock result. In principle, we could expand the series to higher and higher order to improve upon the Hartree-Fock result, but this is infeasible in practice. Furthermore, the expansion diverges for metals. To remedy this problem, Hedin [80] replaced the bare Coulomb potential v with a screened Coulomb potential W. Since W is much weaker than v, it allows for a good approximation even after truncating the expansion at the first term. Σ is therefore approximated as

$$\Sigma(\mathbf{r}, \mathbf{r}', \omega) = \frac{i}{2\pi} \int_{-\infty}^{\infty} G(\mathbf{r}, \mathbf{r}', \omega - \omega') W(\mathbf{r}, \mathbf{r}', \omega') e^{i\delta\omega'} \, d\omega' \,. \tag{4.7}$$

Symbolically, this can be written as

$$\Sigma = iGW, \tag{4.8}$$

which give rise to the name GW approximation.

We can write a Schrödinger-like equation similar to the KS DFT Schrödinger equation in eq. (2.20) using Σ for the QP

$$\left(-\frac{1}{2}\boldsymbol{\nabla}^2 + V + V_{\mathrm{H}} + \Sigma\right)\psi_n^{\mathrm{QP}} = \epsilon_n \psi_n^{\mathrm{QP}}. \tag{4.9}$$

However, in practical applications, we can assume that the KS wavefunctions are similar to the GW wavefunctions, which is the case for many weakly-correlated to moderately- correlated systems. Hence, by comparing eq. (2.20) and the above Schrödinger equation for QP, we can calculate the QP energies using

$$\epsilon_{n\mathbf{k}}^{\mathrm{QP}} = \epsilon_{n\mathbf{k}}^{\mathrm{DFT}} + \langle n\mathbf{k}|\Sigma(\epsilon_{n\mathbf{k}}^{\mathrm{QP}}) - V_{\mathrm{xc}}|n\mathbf{k}\rangle \,. \tag{4.10}$$

Here, we are expressing the KS non-interacting states as $|n\mathbf{k}\rangle$, with n being the band index and $\mathbf{k}$ is the reciprocal vector for sampling the Brillouin Zone (BZ) from the DFT calculation. This means that we are using the DFT wavefunctions and energies as a starting point to a GW calculation. It is also possible to repeat the calculation of the second term using the GW energies and wavefunctions, but diagonalising $\langle n\mathbf{k}|\Sigma|n'\mathbf{k}\rangle$ instead, to improve the QP energies, but this is often not necessary as it is simpler to improve the DFT starting point or use approximation techniques such as the effective energy technique [84].

4.3 Screened Coulomb Potential

We now turn our attention to the screened Coulomb potential which is defined as

$$W(\mathbf{r}, \mathbf{r}', \omega) = \int \epsilon^{-1}(\mathbf{r}, \mathbf{r}'', \omega) v(\mathbf{r}'', \mathbf{r}') \, d\mathbf{r}'' \,, \tag{4.11}$$

where v is the bare Coulomb potential and ϵ^{-1} is the inverse dielectric function

$$\epsilon^{-1}(\mathbf{r}, \mathbf{r}', \omega) = \delta(\mathbf{r}, \mathbf{r}') + \int v(\mathbf{r}, \mathbf{r}'') \chi(\mathbf{r}'', \mathbf{r}', \omega) \, d\mathbf{r}'' \,, \tag{4.12}$$

with χ being the response function defined as

$$\chi(\mathbf{r}, \mathbf{r}', \omega) = -\frac{i}{2\pi} \int G(\mathbf{r}, \mathbf{r}', \omega') G(\mathbf{r}, \mathbf{r}', \omega' - \omega) \, d\omega' \tag{4.13}$$

within the random phase approximation (RPA) [85]. The RPA assumes that electrons respond only to an external potential that is oscillating at a single frequency, so that the phase difference between the particle response and the wave producing it is independent of the position of the particle. This is justified by the fact that the response to phase difference with the wave producing it which depends on the position of the particle tends to average out to zero for a large number of electrons, as the position of the particles are random.

Notice that the self-energy operator (eq. (4.7)) depends on the screened potential which depends on the GF, which in turn depends on the self-energy operator. Therefore, this forms a set of self-consistent equations. However, instead of performing a full self-consistency calculation, we can once again turn to the Dyson series to compute the screened potential:

$$W = v + v\chi^0 W, \tag{4.14}$$

where the χ^0 is the non-interacting response function. χ^0 can be written similarly to eq. (4.13), but with two non-interacting GF:

$$\chi^0(\mathbf{r}, \mathbf{r}', \omega) = -\frac{i}{2\pi} \int G_0(\mathbf{r}, \mathbf{r}', \omega') G_0(\mathbf{r}, \mathbf{r}', \omega' - \omega) \, d\omega' \,. \tag{4.15}$$

Using the definition of the GF from eq. (4.5) and in the basis of DFT wavefunctions,

χ^0 within the RPA can be written as [83]

$$\begin{aligned}\chi^0(\mathbf{r},\mathbf{r}',\omega) = &\sum_{\text{spin}}\sum_{\mathbf{k}n}^{\text{occ}}\sum_{\mathbf{k}'n'}^{\text{unocc}} \phi^{\text{DFT}*}_{\mathbf{k}n}(\mathbf{r})\phi^{\text{DFT}}_{\mathbf{k}'n'}(\mathbf{r})\phi^{\text{DFT}*}_{\mathbf{k}'n'}(\mathbf{r}')\phi^{\text{DFT}}_{\mathbf{k}n}(\mathbf{r}') \\ &\times\left\{\frac{1}{\omega-\epsilon^{\text{DFT}}_{\mathbf{k}'n'}+\epsilon^{\text{DFT}}_{\mathbf{k}n}+i\delta}-\frac{1}{\omega+\epsilon^{\text{DFT}}_{\mathbf{k}'n'}-\epsilon^{\text{DFT}}_{\mathbf{k}n}-i\delta}\right\}.\end{aligned} \tag{4.16}$$

We can, therefore, calculate the screened potential with only the non-interacting response function starting from DFT wavefunctions and energies.

4.4 The Plasmon-Pole Approximation

By rearranging eq. (4.14), along with the definition of the screened potential eq. (4.11), we can express the inverse dielectric function as a function of the non-interacting response function:

$$\epsilon^{-1} = \left[1 - v\chi^0\right]^{-1}. \tag{4.17}$$

Although this does not look too daunting, numerically speaking, we are required to perform an inverse of a matrix for many frequencies, which is very computationally expensive. We can, however, take advantage of the fact that the imaginary part of W is usually flat except for a strong peak corresponding to a plasmon excitation at the plasmon frequency. The plasmon-pole approximation (PPA) assumes that all the weight in $\text{Im}\{W\}$ resides in the plasmon excitation [86–88]. ϵ^{-1} can therefore be approximated by a single-pole function in ω, which in its simplest form can be written as $\text{Im}\left\{\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\right\} = A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})\delta(\omega-\omega_q)$.

Hybertsen and Louie [89] first developed the generalised PPA where the real and imaginary parts of ϵ^{-1} are

$$\text{Re}\left\{\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\right\} = \delta_{\mathbf{G},\mathbf{G}'} + \frac{\Omega^2_{\mathbf{G},\mathbf{G}'}(\mathbf{q})}{\omega^2-\tilde{\omega}^2_{\mathbf{G},\mathbf{G}'}(\mathbf{q})} \tag{4.18}$$

and

$$\text{Im}\left\{\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)\right\} = A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})\left[\delta(\omega-\tilde{\omega}_{\mathbf{G},\mathbf{G}'}(\mathbf{q})) - \delta(\omega+\tilde{\omega}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}))\right], \tag{4.19}$$

with

$$\Omega^2_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) = -A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})\tilde{\omega}_{\mathbf{G},\mathbf{G}'}(\mathbf{q}) \tag{4.20}$$

The unknown parameters $A_{\mathbf{G}\mathbf{G}'}(\mathbf{q})$ and $\tilde{\omega}_{\mathbf{G},\mathbf{G}'}(\mathbf{q})$ can be determined by two different methods. In the original Hybertsen-Louie (HL) PPA [89], they are determined by

using the zero-frequency limit (static limit) and the f-sum rule, which is completely parameter-free. On the contrary, the PPA developed by Godby and Needs [90], called the Godby-Needs (GN) PPA, the parameters are found by fitting $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega)$ at two imaginary frequencies ($\omega = 0$ and $\omega = i\omega_p$). It has been shown that GN PPA reproduces the actual $\mathrm{Re}\{\epsilon^{-1}\}$ better than HL PPA [91].

As can be seen by applying the PPA, the original problem of needing to invert a matrix many times to compute the inverse dielectric matrix is reduced to just finding two unknown parameters for each $\mathbf{G}$, $\mathbf{G}'$ pair. Hence, this reduces the computational cost drastically. A drawback to the PPA is that the lifetime of the quasiparticles cannot be calculated because to the imaginary part of the self-energy is zero except at the plasmon poles.

The GWA with PPA has already been implemented into various codes such as Abinit [72], VASP [92], BerkeleyGW [93] and Yambo [94], and it has shown to predict the band gap with much higher accuracy than DFT [95] and it can be used along with the Bethe–Salpeter equation (BSE) [96] to produce accurate optical spectra [97–99]. Chapter 6 will also demonstrate how the GWA can produce accurate band structures that is comparable to experimental measurements.

Chapter 5

Defects in TMDC Lateral Heterostructures

5.1 Introduction

Lateral heterostructures (LHSs) are a type of heterostructures where multiple materials are grown together in-plane, connected by covalent bonds. This chapter will explore the formation of different types of defects in TMDC LHSs using DFT calculations. TMDC LHSs have been studied extensively experimentally, and results have shown that they can be grown with an atomically sharp interface [100, 101], which can generate strong photoluminescence (PL) at the interface, and their optical and electrical properties can be controlled explicitly [102–105]. Hence, it has been suggested that they could open up a new paradigm for next-generation electronic devices.

The results presented in this chapter focus on three types of TMDC: WS_2, MoS_2 and $MoSe_2$, which are each well studied members of the TMDC class. Their structure consists of stacked trilayers of X-M-X atoms with hexagonal symmetry (space group $P6_3/mmc$), where M is the transition metal and X is the chalcogen. These TMDCs are chosen due to their well known monolayer properties, but their LHS form being lesser known. A comprehensive understanding of structural defects is of paramount importance, as having structural defects can be either detrimental or beneficial, depending on the targeted application. Hence, this chapter concentrates on understanding defects in LHSs, and will begin by looking at vacancies in LHSs with small lattice mismatch, then followed by interface misfit dislocations in LHSs with large lattice mismatch.

Due to the construction of the LHSs requiring large number of atoms (ranging

from 400 to 1300 atoms in this chapter), the linear-scaling DFT ONETEP package [53, 61, 70], the theoretical considerations of which are detailed in chapter 3, was used to perform all the calculations in this chapter. To ensure the convergence of the calculations, test calculations were performed on MoS_2 at varying NGWF radius and kinetic energy cutoff. The values $13\,a_0$ and 1200 eV were found to be sufficiently converged. The PAW formalism described in section 3.5 was employed, where the PAW sets were constructed using the Perdew-Burke-Ernzerhof (PBE) XC functional [33], developed by Garrity et al [106]. Before each of the energy and properties calculation, a geometry optimisation is performed on the system of interest. See table 5.1 for a summary of the important parameters used.

Parameter	Value
Plane-wave cutoff energy	1200 eV
NGWF radius	13 Bohr
XC functional	PBE
Energy change tolerance for NGWF optimisation	$2 \times 10^{-5}\,E_\mathrm{h}$
Convergence threshold for NGWF RMS gradient	8×10^{-7}
Force tolerance for geometry optimisation	$0.002\,E_\mathrm{h}\,{a_0}^{-1}$

TABLE 5.1: A table showing several of the most important parameters used in ONETEP, which would be necessary to reproduce the calculations performed in this chapter.

5.2 Sulphur Vacancy Formation Energy of $WS_2|MoS_2$

The first system explored is a LHS with a small lattice mismatch, i.e., the two materials have similar lattice constants. Within the TMDC family, TMDCs having the same chalcogen, but different transition metal (e.g. MoS_2 and WS_2), have very similar lattice constant. Conversely, TMDCs having the same metal, but different chalcogen (e.g. MoS_2 and $MoSe_2$), have very different lattice constant.

We have chosen to study the $WS_2|MoS_2$ LHS. The lattice constant of WS_2 and MoS_2 are 3.158 Å [107] and 3.169 Å [108], respectively, therefore their lattice mismatch is only ∼0.35%. With such a small lattice mismatch, it is possible to strain one material to have the same lattice constant as the other material, because the energy associated with the strain would be smaller than the energy associated with introducing a dislocation to relieve the strain. Note that this is only true up to some width of the system, which will be discussed in a later section.

We have chosen to strain the WS_2 to have the lattice constant of MoS_2 (3.169 Å). This choice is arbitrary as we believe that straining either material would

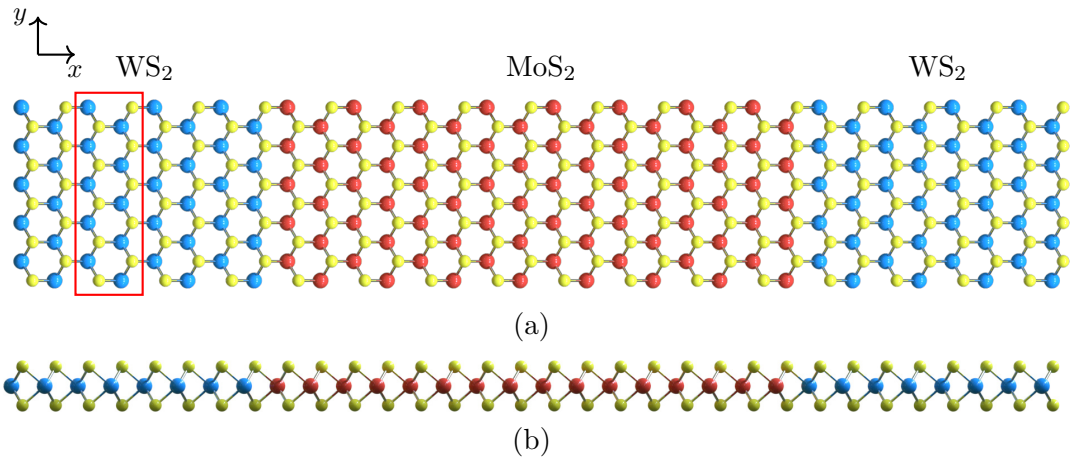


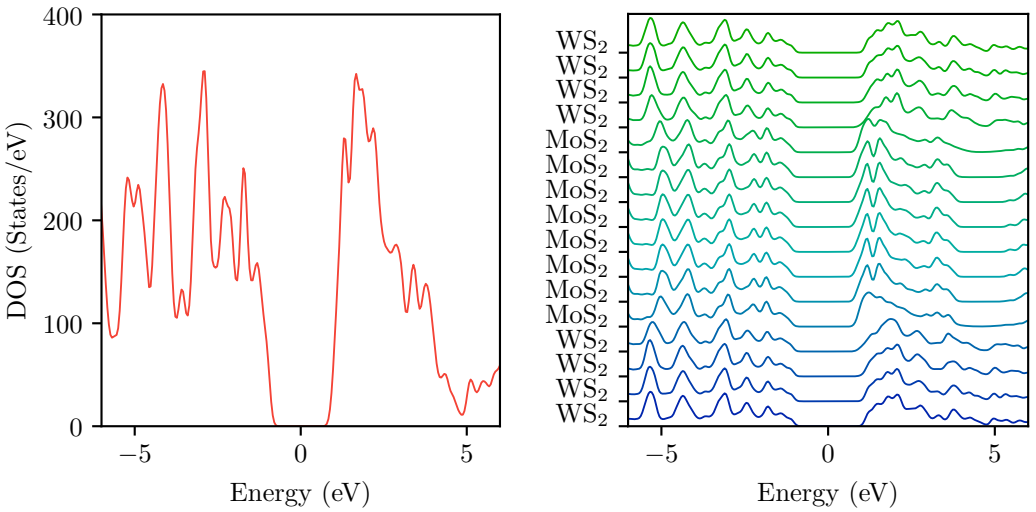
FIGURE 5.1: Schematic of the $WS_2|MoS_2$ LHS with the interface along the zigzag direction, where the blue atoms are W, the red atoms are Mo and the yellow atoms are S. The area defined by the red box is a column or a slab for the slabwise local density of states calculation. (a) is the top view showing a hexagonal structure throughout the LHS and (b) is the side view showing a stacked trilayer of sulphur—transition metal—sulphur.

yield very similar results. A TMDC LHS can be constructed in either the armchair or the zigzag direction, but it has been found experimentally that a zigzag interface is more commonly observed than an armchair interface [103]. Hence, we have constructed the $WS_2|MoS_2$ LHS with an interface along the zigzag direction containing 480 atoms (see fig. 5.1). The size of the cell in non-periodic direction is 14 Å, therefore our system have a vacuum of 11 Å, which is enough to remove any spurious interactions between the periodic images.

5.2.1 Density of States

With the $WS_2|MoS_2$ LHS set up, the total density of states (DOS) was first examined and it is shown in fig. 5.2a. It is of no surprise that fig. 5.2a shows a clear direct band gap. As proved previously, although TMDC in bulk is only semiconducting with an indirect band gap, its monolayer counterpart possesses a direct band gap [109]. The band gap was calculated to be 1.8 eV which is in agreement with previous theoretically results [103].

The slabwise local density of states (LDOS) was then examined. A slab LDOS is defined as follows: the monolayer is divided uniformly into 16 slabs in the x direction, see fig. 5.1a. The slab LDOS is the sum of the contributions to the total DOS from the local orbitals centered on those atoms [110, 111]. Superposing these slab LDOS, as in fig. 5.2b, shows the transition of the DOS from WS_2 to MoS_2.



(a) (b)

FIGURE 5.2: The density of states of $WS_2|MoS_2$, where (a) is the total density of states and (b) is the slabwise LDOS.

The LDOS result shows that the band alignment is a type II heterojunction (staggered gap), similar to the results reported in literature [102, 112]. As experimental study [103] showed that $WS_2|MoS_2$ LHS can generate strong PL at the lateral interface, one might expect a large band gap mismatch at the interface. On the contrary, Figure 5.2b shows that the shift is minimal at the interface. The transition from bulk-like DOS of WS_2 to bulk-like DOS of MoS_2 takes place over just 1–2 slab widths. It is therefore conjectured that the strong PL response could be an effect of separation of defects. Hence, we studied the defect formation energy to gain understanding on where defects are most likely to be observed.

5.2.2 Zhang-Northrup Formalism for Defect Formation Energy

Before we delve into our $WS_2|MoS_2$ vacancy formation energy results, this section will first describe the theory behind the calculation of defect formation energy. In the Zhang-Northrup formalism [113], the formation energy ΔE_f is

$$\Delta E_f = E_T^{\mathrm{def},q} - E_T^{\mathrm{perf}} - \sum_i \Delta n_i \mu_i + q\mu_e, \tag{5.1}$$

where $E_T^{\mathrm{def},q}$ and E_T^{perf} are the total energies of the defect and perfect crystals, μ_i is the chemical potential of the atomic species i, Δn_i is the difference in the number

of atoms of the element i to create the defect, and μ_e is the chemical potential of an electron in the perfect crystal. Note that since ΔE_f is the energy associated with creating a defect, the lower it is, the higher the concentration of defects.

For the specific case of $WS_2|MoS_2$ with a single neutral S vacancy defect, the system of interest here, we can write

$$\Delta E_f = E_T^{\text{def}} - E_T^{\text{perf}} + \mu_{\text{S}}, \tag{5.2}$$

where μ_{S} is the chemical potential of an S atom. This chemical potential is a rather troublesome quantity to define, as it is highly dependent on temperature and pressure for a compound. This quantity is thus usually bounded by the experimental conditions. We can define upper and lower bounds in which the chemical potential must lie, since the formation energy can be extracted as a range of values for different formation conditions.

To obtain the bounds of μ_{S}, let us examine the specific case of $WS_2|MoS_2$. From the fact that it could be imagined to be formed exothermally from bulk W, bulk Mo and gaseous S, one can deduce that the chemical potential per atom of both species in the monolayer must not be higher than it would be in their elemental states. Hence, the following inequalities can be written:

$$\mu_{\text{W}} \leq \mu_{\text{W}}^{\text{bulk}}, \quad \mu_{\text{Mo}} \leq \mu_{\text{Mo}}^{\text{bulk}} \quad \text{and} \quad \mu_{\text{S}} \leq \mu_{\text{S}}^{\text{gas}}, \tag{5.3}$$

where $\mu_{\text{W}}^{\text{bulk}}$, $\mu_{\text{Mo}}^{\text{bulk}}$ and $\mu_{\text{S}}^{\text{gas}}$ are the chemical potentials for bulk Mo, bulk W and gaseous S, respectively.

In order to find the lower bound, consider the Gibbs free energy of formation of the $WS_2|MoS_2$,

$$\Delta G_f^{\text{WS}_2|\text{MoS}_2} = \mu_{\text{WS}_2|\text{MoS}_2} - \frac{1}{2}\mu_{\text{W}}^{\text{bulk}} - \frac{1}{2}\mu_{\text{Mo}}^{\text{bulk}} - 2\mu_{\text{S}}^{\text{gas}} \tag{5.4}$$

where $\mu_{\text{WS}_2|\text{MoS}_2}$ is the total Gibbs free energy per formula unit of the compound and can be expressed as the sum of the chemical potential of each element:

$$\mu_{\text{WS}_2|\text{MoS}_2} = \frac{1}{2}\mu_{\text{W}} + \frac{1}{2}\mu_{\text{Mo}} + 2\mu_{\text{S}}. \tag{5.5}$$

From the above definition, we can establish the fact that μ_{S} is at its minimum only when μ_{W} and μ_{Mo} are at their maximum, i.e., at their bulk form: $\max(\mu_{\text{W}}) = \mu_{\text{W}}^{\text{bulk}}$ and $\max(\mu_{\text{Mo}}) = \mu_{\text{Mo}}^{\text{bulk}}$. Using this knowledge and the three preceding equations,

we can determine the lower bound of μ_S, which is given by

$$\min(\mu_S) = \frac{1}{2}\Delta G_f^{WS_2} + \mu_S^{gas}. \tag{5.6}$$

And the upper bound is simply

$$\max(\mu_S) = \mu_S^{gas}. \tag{5.7}$$

The upper and lower bounds of μ_S shall be known as S-rich and S-poor henceforth.

5.2.3 Vacancy Formation Energy

Since we believe that the strong PL response at the interface could be an effect of defects, we turn our attention to investigating how vacancies are distributed on a LHS. We have decided to investigate vacancies because they are both the simple to create and are the most likely to be observed in experiment. In the case of $WS_2|MoS_2$ LHS, we can introduce W, Mo, single S or double S vacancies. However, the absence of a transition metal atom or double S would be quite rare. It was found that the vacancy formation energy of a single S is the lowest among the above types of vacancy [114]. Hence, the choice of analysing the single S vacancy formation energy for this project.

The theoretical framework described in the previous subsection suggests that calculating the vacancy formation energies for $WS_2|MoS_2$ requires the simulations of bulk W, bulk Mo, gaseous S and $WS_2|MoS_2$ with and without an S vacancy. The bulk W and Mo models were set up to be a $4 \times 4 \times 4$ supercell of the body-centered cubic structure with the lattice parameter 3.192 Å for W [115] and 3.169 Å for Mo [115]. The gaseous S was set up to be cyclo-S_8 with a crown-like structure in a $20 \times 20 \times 20$ Å simulation cell, where the bond length is 2.07 Å obtained from geometry optimisation. Since we are interested in where an S vacancy is most likely to form on the heterostructure, we set up 16 different calculations of a $WS_2|MoS_2$ with a single top S missing at different columns. Figure 5.3 shows the location of where each S vacancy is placed. This will give us the vacancy formation energy as a function of position across the LHS. In addition, we have calculated the vacancy formation energy of WS_2 and MoS_2 monolayers S as reference. They were calculated using a 5 row by 3 column monolayer with and without an S vacancy with the lattice parameter 3.169 Å [108] (see fig. 5.4).

The total energy of each of the bulk systems described above is calculated and divided by their total number of formula unit to get their respective chemical potential. The results for the calculations are shown on fig. 5.5. As can be seen

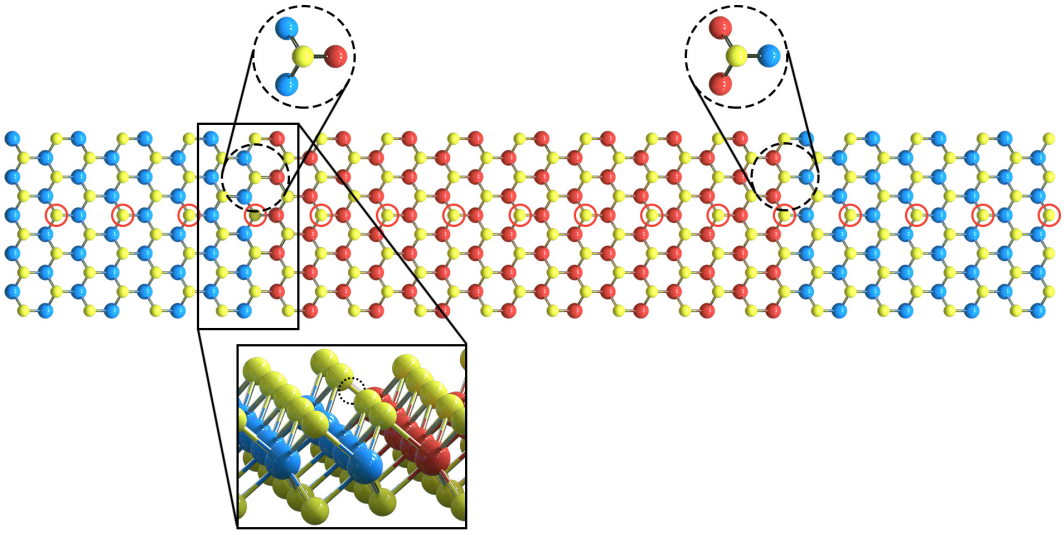
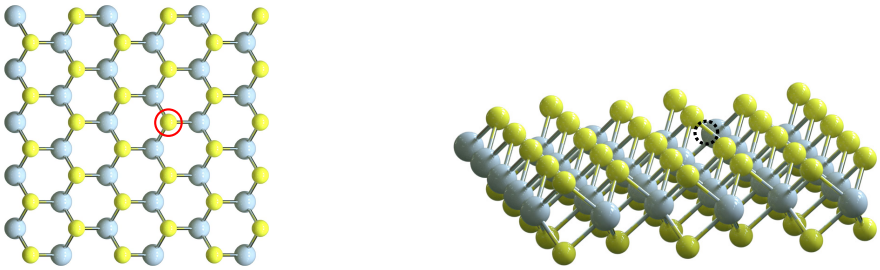
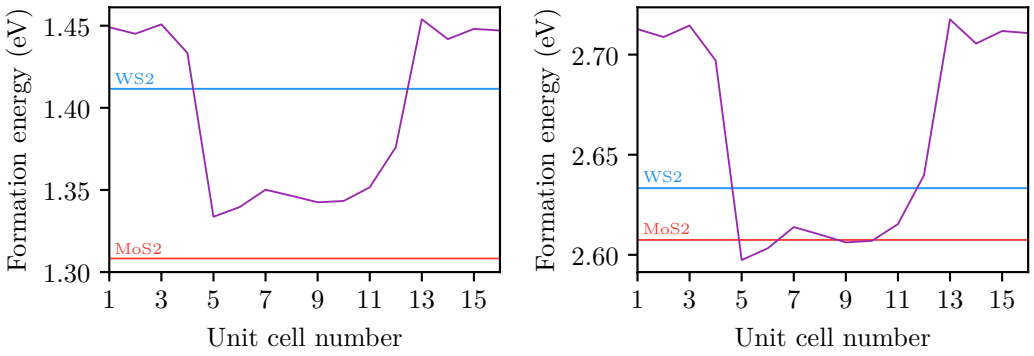


FIGURE 5.3: Schematic of the $WS_2|MoS_2$ with red circles showing where each S atom is removed along the HS. The bottom zoomed-in inset highlights the missing top S atom. The top two zoomed-in insets highlights the two inequivalent interfaces, in which the left interface is built with the S atoms connected to two W atoms and one Mo atom, whereas the right interface is built with the S atoms connected to two Mo atoms and one W atom.



(a) (b)

FIGURE 5.4: Schematic of a 5 rows by 3 columns MX_2 supercell, where the grey atoms are either W or Mo and the yellow atoms are S. (a) is the top view with the red circle showing where the S vacancy is placed and (b) is the side view highlighting that the S vacancy only appears on the top layer.

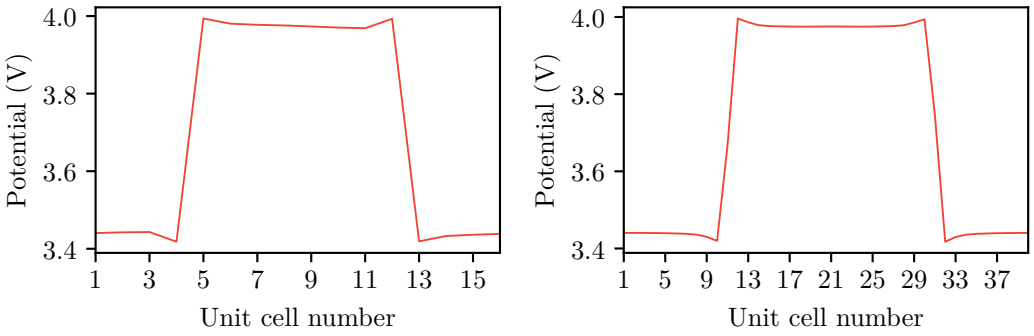


(a) S-poor (b) S-rich

FIGURE 5.5: Vacancy formation energy of the $WS_2|MoS_2$ LHS as a function of position, where (a) is for the S-poor case and (b) is for the S-rich case which defines the range of possible values for the formation energy. The red line and blue line are the S vacancy formation energy of MoS_2 and WS_2, respectively.

that the S vacancies have lower formation energy in the MoS_2 region than in the WS_2, thus S vacancies are more likely to form in the MoS_2 region. This is apparent from the vacancy formation of their separate monolayer counterpart, also marked on fig. 5.5, because the vacancy formation energy of WS_2 is higher than MoS_2. The LHS also increased the overall vacancy formation energy when compared to their separate monolayer counterpart. In addition, a notable feature of the results is that there is a slight tendency for vacancies to segregate to the W|Mo interface (left inset of fig. 5.3) rather than the Mo|W interface (right inset of fig. 5.3). This slight asymmetry was investigated by looking at the average electric potential for each column and was found that the asymmetry also exists in the electric potential, see fig. 5.6a. We hypothesise that the trapping of vacancies near one of the interfaces, albeit only a small effect, could form part of the reason for enhanced PL.

If we look closely at the $WS_2|MoS_2$ LHS, one can see that the zigzag interfaces on the left and on the right are in fact different. Figure 5.3 highlights the differences, where on the left side, the S atoms are connected to two W atoms and one Mo atom, whereas on the right side the S atoms are connected to two Mo atoms and one W atom. These inequivalent interfaces cause a small electric field across the system, therefore creating an asymmetric electric potential and hence the asymmetry of the vacancy formation energy. To see that this effect is in fact due to the interfaces, we looked at the same LHS but with interfaces going along the armchair direction, as illustrated in fig. 5.7. As we can see from fig. 5.7, the two interfaces are now equivalent. The average electric potential for such system is shown in fig. 5.6b and



(a) Zigzag interface (b) Armchair interface

FIGURE 5.6: Average electric potential for $WS_2|MoS_2$ LHS across the whole structure. The zigzag interface in (a) illustrates an asymmetric potential as shown by a small downward slope across unit cell number 6 and 11, whereas the armchair interface in (b) shows a symmetric potential.

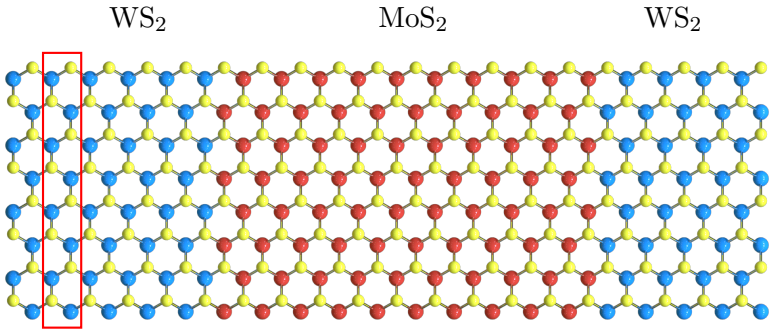


FIGURE 5.7: Schematic of the $WS_2|MoS_2$ LHS with the interface along the armchair direction, where the blue atoms are W, the red atoms are Mo and the yellow atoms are S. The area defined by the red box is a column or a slab for generating the average potential.

it shows a symmetric potential, hence proving that two inequivalent zigzag interfaces can create a small dipole field.

This concludes our study of vacancies in LHS with small lattice mismatch. The next three sections will be the discussion on defects in LHS with large lattice mismatch.

5.3 Critical Thickness and Dislocation

For LHSs with a small lattice mismatch, we argued that it is possible to simply strain one material such that their lattice constants match. Experimentally, a LHS can be described as consisting of a 2D film and a 2D substrate (see fig. 5.8). The 2D

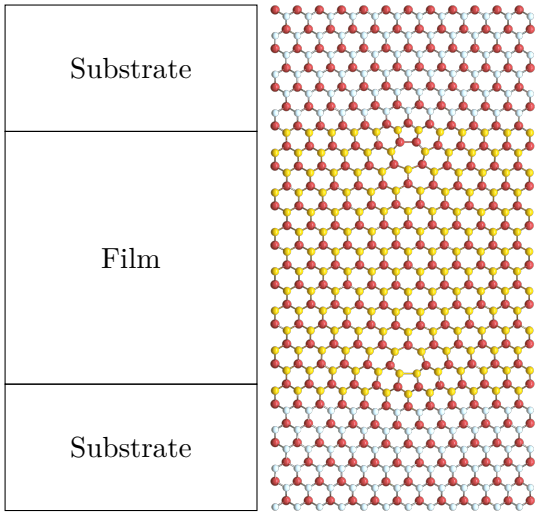
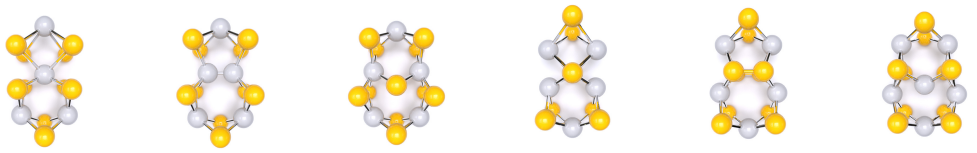


FIGURE 5.8: Schematic of a periodic cells with a 2D substrate and a 2D film.

substrate would be the first deposited material, and the 2D film would be expected grow from the island edge of this 2D substrate. We shall henceforth denote the 2D substrate and 2D film as simply substrate and film to be understood as referring to in-plane growth; and the reader should not confuse these terminologies with their usual definitions. See fig. 5.8 on how a simulation cell can be set up according to our terminologies.

As the film grows and the film thickness increases, the interfacial strain gradually builds up. At the so-called critical thickness, it becomes energetically favorable to relieve the strain by including some structural transformation in the film. The critical thickness of small lattice mismatch LHSs, however, would be very high. For example, a hexagonal boron nitride (h-BN) film grown on a graphene substrate with ∼2.4% mismatch, which is already a relatively large mismatch, the critical thickness is 30 Å [116], more than 10 h-BN unit cells. One can imagine that a $WS_2|MoS_2$ LHS with only ∼0.35% mismatch will have a much larger critical thickness, thus straining the film is more likely over fairly long distance.

The critical thickness would be very small for large lattice mismatch LHSs, hence becomes a very important and unavoidable parameter. The structural transformation required to relieve the strain is a network of misfit dislocations at the interface, a type of defects which this section will focus on. Since TMDC LHSs usually have a zigzag interface, the strain could be relaxed through an insertion or removal of a column of atoms perpendicular to the interface, in the armchair direction. Similar to graphene [117], when a column of armchair atoms are removed and the resulting dangling bonds are reconnected, the pentagon-heptagon (5|7) disloca-



(a) 4|6 ⊥ (b) 5|7 ⊥ (c) 6|8 ⊥ (d) 4|6 ⊤ (e) 5|7 ⊤ (f) 6|8 ⊤

FIGURE 5.9: Three common types of edge dislocation for 2D TMDCs: 4|6, 5|7 and 6|8 dislocations in their ⊥ and ⊤ form. The grey atoms are transition metal and the orange atoms are chalcogen. Notice that the ⊤ dislocations are simply the ⊥ dislocations with their metal and chalcogen atoms interchanged.

tion core naturally appears. Although unlike graphene, TMDCs (MX_2) can create two different types of dislocation: an M-rich dislocation with a metal homoelemental bond corresponding to Burgers vector $\mathbf{b} = (1, 0)$ and an X-rich dislocation with chalcogen homoelemental bonds corresponding to Burgers vector $\mathbf{b} = (0, 1)$. The metal-rich dislocations are usually identified as ⊥ and the dichalcogenide-rich dislocations are usually identified as ⊤.

Aside from the 5|7 dislocation commonly seen in 2D TMDCs, there are two other dislocations that are very likely to form: the rhombus-hexagon (4|6) dislocation and the hexagon-octagon (6|8) dislocation [118]. Both of these dislocations exists in ⊥ form and ⊤ form. The structures of all the dislocations cores mentioned are shown in fig. 5.9. The likelihood of one type of dislocation forming is dependent on the chemical potential of the environment, where the 4|6 dislocation is most likely to form under metal-rich environment and the 6|8 dislocation is most likely to form under S-rich environment [118, 119]. However, note that it has been found that chemical vapour deposition grown MoS_2 film can have both 4|6 and 6|8 dislocation regions due to local fluctuations of Mo and S source [119]. Since the 4|6 and 6|8 dislocations can be viewed as the derivative of the 5|7 dislocation by removing or adding an atom at the core, we have decided to concentrate on the 5|7 dislocation. With the critical thickness and the basic theory of dislocation defined, the next section will detail a simple approximation to calculating the critical thickness using first-principle methods.

5.4 Simple Approach to Approximating the Critical Thickness of $MoS_2|MoSe_2$ using DFT

We will consider the simple case of a thin film on a thick substrate. That is, the film is much smaller than the substrate. In this case, only the film will be strained, so one

simply needs to be able to calculate the strain as a function of the film thickness and the energy associated with creating a dislocation in order to calculate the critical thickness. In this section we will study two configurations of $MoS_2|MoSe_2$ LHS, where one is with MoS_2 as the substrate and the other one is with MoS_2 as the film.

A 5 rows by 3 columns supercell, same as the one shown in fig. 5.4, was used to compute the strain energy per formula unit, which can then be used to calculate the strain as a function of the film thickness. The energy per formula unit we refer to here is simply the total energy divided by the number of formula units in our simulation cell. For example, for a simulation cell containing 315 atoms for MoS_2 (three atoms, one Mo and two S, in each formula unit), the energy per formula unit will be the total energy of the system divided by 105.

For both MoS_2 and $MoSe_2$ the total energy of a strained cell, where the cell size in the zigzag direction is set to be the same as the substrate, and a perfect cell were calculated. The strain energy per formula unit is then calculated by the energy difference between these two cells divided by the number of formula units:

$$E_{\text{strain}} = E_{\text{strained}} - E_{\text{perfect}}. \tag{5.8}$$

Note that the strain is only applied in the zigzag direction because the strain caused by the substrate is only in the interfacial direction.

As for extracting the dislocation core energy, we approximate this by finding the difference between the energy of a nanoribbon with a dislocation and a perfect nanoribbon:

$$E_{\text{disloc}} = E_{\text{NR}}^{\text{disloc}} - E_{\text{NR}}^{\text{perfect}}. \tag{5.9}$$

Following the Burgers model, we have created a nanoribbon with a low-angle grain boundary containing a dislocation [42, 120]. We will first present the geometry of the grain boundary and the rotated nanoribbon.

According to the Burgers model [120], the interspacing between a pair of adjacent dislocation cores depends on tilt angle θ. From fig. 5.10, one can see that

$$h = \frac{a}{2\sin\left(\frac{\theta}{2}\right)} \approx \frac{a}{\theta}. \tag{5.10}$$

Because of the geometry of the 5|7 dislocation, the height of the dislocation is approximately $3\sqrt{3}a/2$. This is also the same with the 4|6 and 6|8 dislocations. Since the height of a perfect unit cell is $\sqrt{3}a$, one can generate different dislocation cores distances by using different number of perfect cells between two adjacent dislocation

cores, this means that we can write

$$h_n = \left(n + \frac{1}{2}\right)\sqrt{3}a + \Delta, \tag{5.11}$$

where n is an integer and Δ is a very small variable which is described later. From eq. (5.10) and eq. (5.11), we can deduce the misorientation angle to be

$$\theta_n \approx \frac{1}{\sqrt{3}\left(n + \frac{1}{2}\right)}. \tag{5.12}$$

With this angle, a grain boundary can be constructed by rotating two normal nanoribbons, where one is rotated by angle $-\frac{\theta}{2}$ and the other rotated by angle $\frac{\theta}{2}$, then seamlessly connect them in the middle, which will result in fig. 5.10.

We will now calculate the value of Δ by looking at the simpler case of a rotated nanoribbon. If one uses a normal nanoribbon then rotates it by some angle while keeping the same cell size, their periodic images would overlap. This behaviour is shown on fig. 5.11. If the height of the cell of the normal nanoribbon is h and the height of the overlap region is Δ, then the height of the new cell is

$$h_{\text{rotated}} = h + \Delta. \tag{5.13}$$

To obtain Δ, let us consider the rotation of the point at the center of fig. 5.11 in between the two periodic cells. After a rotation of angle θ, the point would have been rotated to the left and right for the bottom cell and the top cell, respectively, giving

$$\mathbf{P}_a = \begin{pmatrix} -\frac{h}{2}\sin(\theta) \\ \frac{h}{2}\cos(\theta) \end{pmatrix} \quad \text{and} \quad \mathbf{P}_b = \begin{pmatrix} \frac{h}{2}\sin(\theta) \\ h - \frac{h}{2}\cos(\theta) \end{pmatrix}. \tag{5.14}$$

Since Δ is just the change in the y-axis, it is simply

$$\Delta = h(1 - \cos(\theta)). \tag{5.15}$$

therefore the height of the rotated nanoribbon is

$$h_{\text{rotated}} = h(2 - \cos(\theta)). \tag{5.16}$$

In the case where θ is given by eq. (5.12), the height of a grain boundary nanoribbon or a rotated nanoribbon is

$$h_n = \left(n + \frac{1}{2}\right)\sqrt{3}a\left(2 - \cos\left(\frac{1}{\sqrt{3}(n + \frac{1}{2})}\right)\right). \tag{5.17}$$

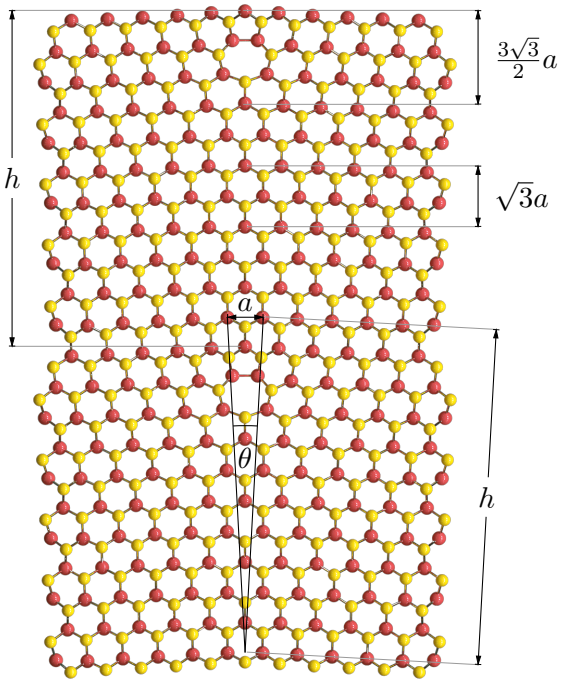


FIGURE 5.10: Top view of a grain boundary constructed by a periodic arrangement with a 5|7 dislocation core. The height of the dislocation core is approximated to be $\sim \frac{3\sqrt{3}}{2}a$ and the height of a unit cell is $\sqrt{3}a$. h is the height of the individual periodic cell. The misorientation angle is labeled as θ.

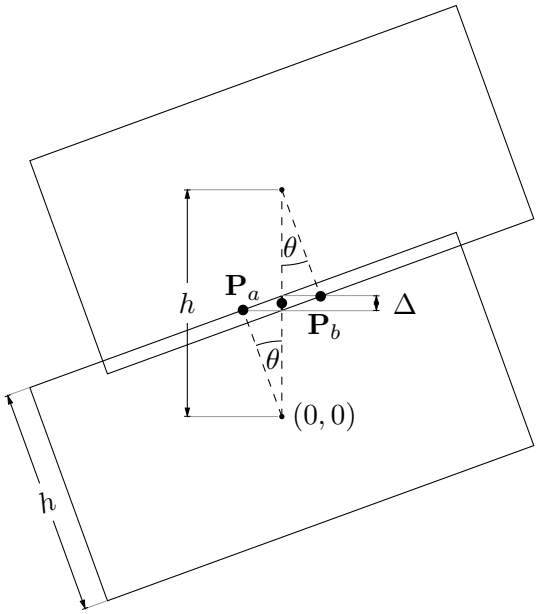
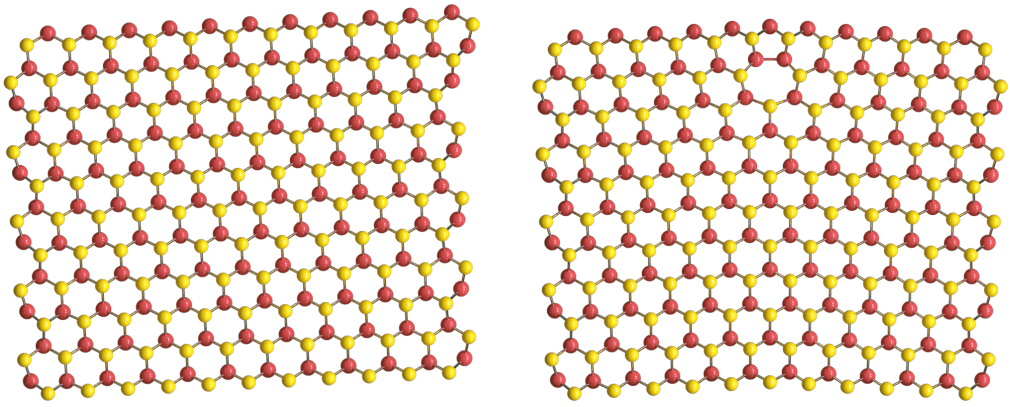


FIGURE 5.11: Schematic of two cells of height h rotated by angle θ about their cell center with an overlap region where the height of this region is Δ.



(a) (b)

FIGURE 5.12: Nanoribbons for calculating the dislocation energy, where (a) is a perfect rotated nanoribbon and (b) is a nanoribbon with a grain boundary in the middle containing a 5|7 dislocation.

$n = 5$ was chosen to create our nanoribbon, which gives us the misorientation angle $\theta = 6°$. This value of n was chosen so that the dislocation core distance is relatively large to remove any spurious interactions between them. Because of the misorientation angle, our perfect nanoribbon is also rotated to cancel out any edge effects. The perfect nanoribbon and the grain boundary constructed are shown in fig. 5.12. To prevent their periodic images from interacting, 10 Å vacuum spacing is used in the left and right direction, and the out-of-plane direction.

We have applied the same approach as ref. [116] to calculate the critical thickness, namely that we compare the total energy of a purely strained film without a dislocation with the total energy of a film that contains a 5|7 dislocation core, embedded at different depths in the film. With the strain energy per formula unit and the dislocation core energy for both MoS_2 and $MoSe_2$, we can construct such a hypothetical system so that we have different film thickness with a dislocation embedded. Our film will contain strained cells before the dislocation and perfect cells beyond the dislocation, thus the energy can be written as

$$E_{\text{film}} = nE_{\text{perfect}}(t_{\text{film}} - d_{\text{disloc}}) + nE_{\text{strain}}d_{\text{disloc}} + E_{\text{disloc}}\,. \tag{5.18}$$

E_{perfect} is the energy per formula unit of a perfect cell; E_{strain} is the energy cost of adding strain to the system (see eq. (5.8)); E_{disloc} is the additional energy cost associated with creating a dislocation (see eq. (5.9)); t_{film} and d_{disloc} are the film thickness in number of unit cells and dislocation distance away from the interface in number of unit cells; n the expected average misfit dislocation spacing in the

interfacial directions.

To calculate n, we consider that the inclusion of a dislocation removes one armchair chain of atoms in the film, while the substrate is intact, thus we can find the dislocation spacing by solving

$$a_{\text{substrate}} n = a_{\text{film}}(n-1), \tag{5.19}$$

where $a_{\text{substrate}}$ and a_{film} is the lattice constant of the substrate and the film, and n is the number of unit cells in the interfacial direction which needs to be found. The lattice constants 3.169 Å [108] and 3.288 Å [121] were used for MoS_2 and $MoSe_2$, respectively, giving $n = 27$ unit cells or interfacial length ~ 85.6 Å. Linear-scaling DFT calculations using such a high value of unit cells would be too computationally expensive, which motivates the development of this approximation.

The results of the embedded dislocation core analysis are shown in fig. 5.13. In fig. 5.13a, we consider a 5|7 dislocation core, embedded at depth d_{disloc} in a MoS_2 film on a thick $MoSe_2$ substrate. In fig. 5.13c, we consider a 5|7 dislocation core in a $MoSe_2$ film on a thick MoS_2 substrate. It is not surprising to see that it is energetically favourable for the dislocation to be at the interface, because the additional strained part of film does not exist, thus no extra strain energy is added to the system. The negative values of figs. 5.13a and 5.13c show when it is energetically more favourable for a dislocation to form than an equivalent sized purely strained supercell to form. The zero crossings for these curves therefore show the maximum embedded core distance at which the dislocation can still provide energy reduction. The results for these zero crossings are shown in figs. 5.13b and 5.13d, and the critical thickness is just the zero crossing of these. The critical thickness for a MoS_2 film on an infinite $MoSe_2$ substrate is 20.97 Å, while the critical thickness for a $MoSe_2$ film on an infinite MoS_2 substrate is 6.61 Å. This large difference is attributed to the fact that straining $MoSe_2$ requires more energy than straining MoS_2.

Although we have opted for this embedded dislocation core analysis to demonstrate the behaviour of a dislocation moving away from an interface and computing the critical thickness from that, it is useful to note that using the energies from eq. (5.18), one can obtain the critical thickness by solving $E_{\text{film}} - E_{\text{strain}} = 0$ for $d_{\text{disloc}} = 0$, giving

$$t_{\text{crit}} = \frac{\sqrt{3}}{2} a_{\text{film}} \frac{E_{\text{disloc}}}{n(E_{\text{strain}} - E_{\text{perfect}})}. \tag{5.20}$$

This equation gives the same critical thickness as fig. 5.13.

Our approach is a simple approximation, nevertheless it seems to fall in the same ballpark as [116] for h-BN|graphene LHS. Since $MoSe_2$ and MoS_2 have a

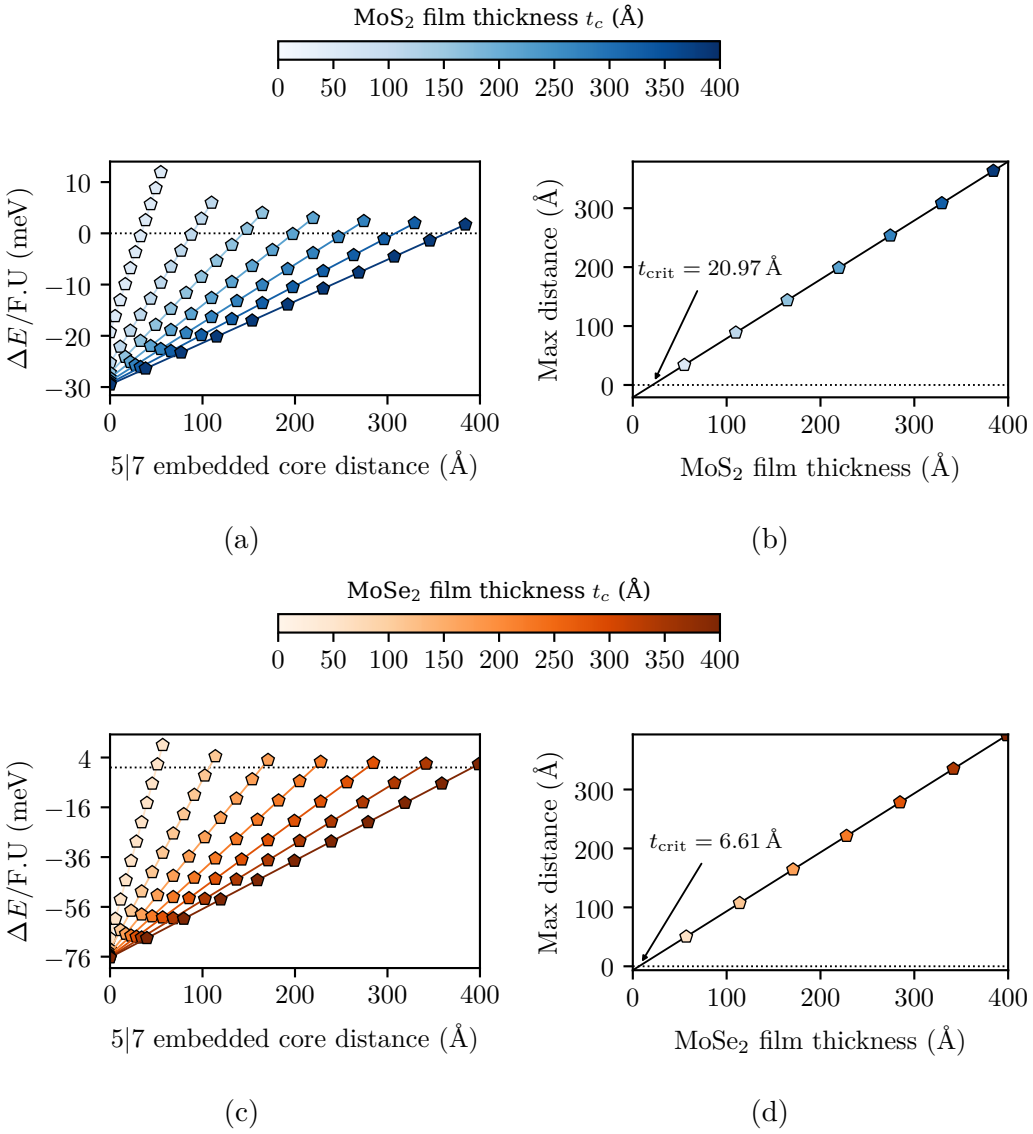


FIGURE 5.13: Embedded dislocation core analysis of (a) MoS_2 film with a 5|7 dislocation on a thick $MoSe_2$ substrate and (c) $MoSe_2$ film with a 5|7 dislocation on a thick MoS_2 substrate. ΔE/F.U is the energy difference between a film with a dislocation and a purely strained film. The lines in (a) and (c) show this energy difference as the dislocation move away from the interface for different film thicknesses. (b) and (d) shows the maximum embedded core distance for energy reduction which are obtained from the zero crossings of (a) and (c). (b) shows the thermodynamic critical thickness for a MoS_2 film on a thick $MoSe_2$ substrate is interpolated as 20.97 Å. (d) shows the thermodynamic critical thickness for a $MoSe_2$ film on a thick MoS_2 substrate is interpolated as 6.61 Å.

larger lattice mismatch than h-BN and graphene, it is expected that our critical thickness values will be smaller than the critical thickness of h-BN|graphene. This approximation does have two major drawbacks: it fails to describe the energy of a dislocation core correctly, and it completely neglects the strain at the interface. We shall see in the next section that a dislocation in a LHS does not create a grain boundary, it appears that our choice of using a grain boundary to extract the energy of the dislocation greatly reduces the real local strain field caused by the dislocation and hence the energy associated with it. In addition, the strain field goes across both the film and the substrate, an effect which is neglected completely. The next section will discuss the construction of a LHS for first-principle calculations and hence visualise the sort of strain neglected in our simple approximation in this section.

5.5 Constructing realistic lateral heterostructure for first-principle calculations

To construct a LHS with periodic boundary condition, such that it seamlessly connects with its in-plane periodic images, one cannot use just one dislocation core, because after the insertion of the dislocation core in the film has removed one column of atoms, there would not be enough columns of atoms to bond with the columns of atoms in the periodic image of the substrate. This behaviour is illustrated in fig. 5.12, the number of chalcogen pairs at the top of the nanoribbon in is 12, whereas at the bottom is 11. If the substrate is first placed at the top of the nanoribbon, there will be 12 pairs of covalent bonds between the metal atoms and the chalcogen atoms, whereas the bottom has only 11, so it will be missing one more pair of chalcogen in order to covalently bond with every metal atom of the substrate periodic image. It could be confusing that fig. 5.12 shows only 11 metal atoms at the top, but that is because it is not periodic in the left and right direction. An additional metal atom at the top is required to create a model with full 2D periodicity.

We must use two dislocation cores of opposite polarity (i.e., both $\perp$ and $\top$ from fig. 5.9), one on each end of the film. Figure 5.14 illustrates a 5|7 $\perp$ dislocation at the top of the film and a 5|7 $\top$ dislocation at the bottom of the film. By placing a $\top$ dislocation at the bottom, an extra column is reinserted before it reconnects back with the substrate. While one could simply use a perfect cell and add the $\perp$ and $\top$ dislocations accordingly, then perform a geometry optimisation to obtain the optimal structure, this would be unwise as the starting structure would be too far

from optimal. Since a LHS often contains large number of atoms, *ab-initio* geometry optimisation would be too computationally expensive.

In order to put the initial structure closer to the optimal structure, one can utilise the rotation angle created by the dislocation from eq. (5.12). We choose the height of the film similar to that of the design of the grain boundary nanoribbon, then find an angle associated with it. Rather than inserting a grain boundary nanoribbon as the film, we slowly rotate the columns back to 0° when it reaches the edge of the cell. This process is highlighted on fig. 5.14 as blue lines. The change of the angle on each column is just θ/n_{col}, where n_{col} is the number of columns.

From fig. 5.14, it is clear that the simple approximation from the last section has ignored several important strain effects from the LHS. The slow rotation procedure described above would mean that the strain field around the dislocation is much larger than that of a grain boundary. In addition, it can be see that the interface between the two materials is not 'straight', thus there is also strain associated with the interface, which is completely neglected in the simple approximation. One could use a model of the full LHS to compute the critical thickness, but as we will show, this is restrictively computationally expensive. The LHS in fig. 5.14 contains 903 atoms, but it has only 11 columns (or $n = 11$ in eq. (5.19)) and the height of film is double the grain boundary nanoribbon from last section. The ideal number of columns is 27, which means that the ideal structure would be close to 3000 atoms. Running a geometry optimisation on such a large system is very expensive even in linear-scaling DFT. However, it would be necessary to do so to capture the strain correctly. Extrapolating the total energy to the ideal structure from several smaller structures is also not possible since eq. (5.19) only has integer solutions, so the strain is not a smooth function of the number of columns. These are the shortcomings we have faced for this section of this project. The design procedure outlined in this section, however, can certainly be used for smaller LHSs, ones that have very large lattice mismatch.

5.6 Conclusions

We have discussed techniques developed using linear-scaling DFT to analyse defects in LHSs. We firstly explored the $WS_2|MoS_2$ LHS, a LHS with small lattice mismatch. We have computed the DOS and slabwise LDOS. The DOS shows a band gap of 1.8 eV. While the LDOS shows the band alignment is a type II heterojunction, the shift is minimal at the interface. Because of the small difference in the band alignment, we believe that strong PL response could be due to the separation of

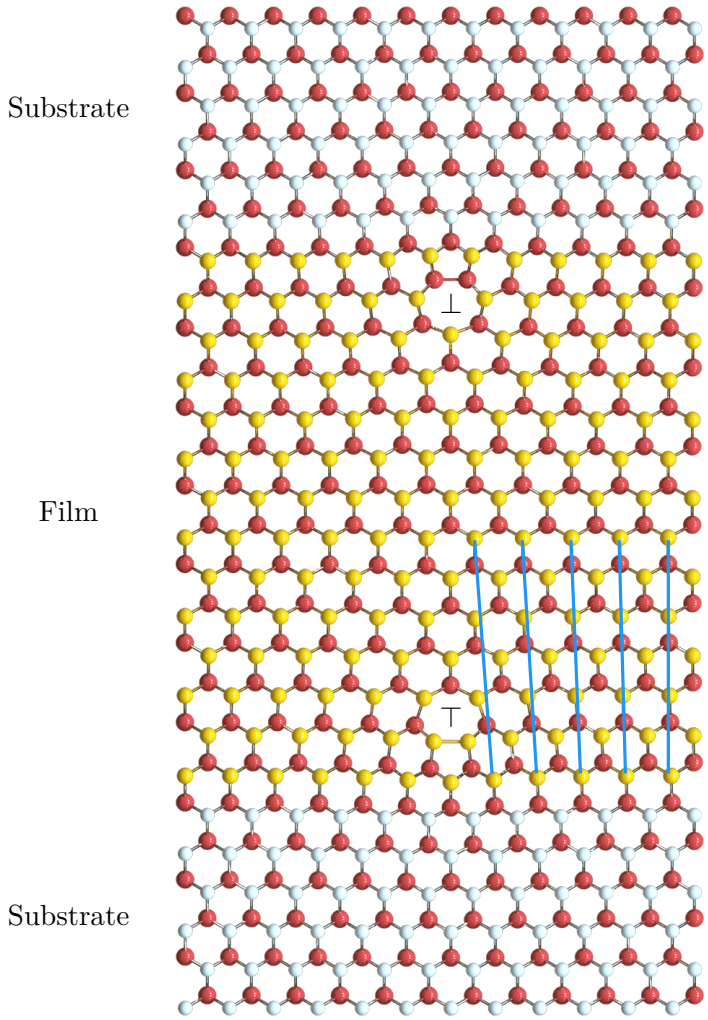


FIGURE 5.14: Schematic of a geometry optimised periodic LHS containing 903 atoms, where we denote the top and bottom regions as substrate and the middle region is the film. The film contains a 5|7 ⊥ dislocation at the top and a 5|7 ⊤ dislocation at the bottom.

defects. We subsequently went on to study how sulphur vacancies are distributed across the LHS.

The vacancy formation energy as a function of position across the LHS was calculated. It was found that sulphur vacancies are more likely to form in the MoS_2 region and with a slight tendency for the vacancies to form at the W-Mo interface. This asymmetry was found to occur only in the zigzag interface. Correspondingly the average potential across the LHS was found to be asymmetric for LHSs with zigzag interfaces, but symmetric for LHSs with armchair interfaces. We hypothesise that the enhanced PL observed experimentally could be due to the trapping of vacancies near one of the interfaces.

We then explored the $MoS_2|MoSe_2$ LHS, a LHS with large lattice mismatch. Because of the dissimilar lattice constants between the two material, dislocations are formed above the critical thickness to relieve the strain. Hence, we have developed a simple approximation to compute the critical thickness of a thin film on a thick substrate, based on DFT calculations. In our model, we used a grain boundary nanoribbon that includes a 5|7 dislocation and a rotated nanoribbon to extract the energy associated with adding a dislocation. Then we obtained the energy per formula unit for a perfect film and a one-directional strained film. Using all the energies calculated, we can construct hypothetical systems that have different film thickness with a dislocation embedded. We obtained the critical thickness of a thin MoS_2 film on a thick $MoSe_2$ substrate as 20.97 Å. As for a thin $MoSe_2$ film on a thick MoS_2 substrate, the critical thickness is 6.61 Å. Despite the crude approximations, the critical thicknesses is in the expected range.

Finally, we have presented how one may design a full periodic LHS to be simulated using DFT. As further work, it would be interesting to use a full periodic LHS to compute the critical thickness. This would be easiest for a LHS with a very large lattice mismatch (e.g., Graphene|MoS_2) to reduce the number of atoms needed.

Chapter 6

Producing Accurate Band Structures of Two-Dimensional Materials Using the GW Approximation

This chapter describes DFT and GW analysis work we have done in collaboration with the University of Washington and the microscopy group at University of Warwick. All of the experimental work presented in this chapter was performed by the experimentalists led by Xiaodong Xu and David H. Cobden, from the University of Washington, and Neil R. Wilson, from the University of Warwick. This work was published in Nature in 2019:

P. V. Nguyen, N. C. Teutsch, N. P. Wilson, J. Kahn, X. Xia, A. J. Graham, V. Kandyba, A. Giampietri, A. Barinov, G. C. Constantinescu, N. Yeung, N. D. M. Hine, X. Xu, D. H. Cobden, and N. R. Wilson, "Visualizing electrostatic gating effects in two-dimensional heterostructures", Nature **572**, 220–223 (2019)

6.1 Introduction

Two-dimensional (2D) van der Waals heterostructures have attracted intense study due to their remarkable electronic properties and applicability in nanoscale devices [4, 15]. If one could directly monitor the states of electrons of these heterostructures as a gate voltage is applied, it could transform our understanding of the physics and the electronic properties of these material, and hence would aid the discovery of new electronic devices. Here we demonstrate this ability by

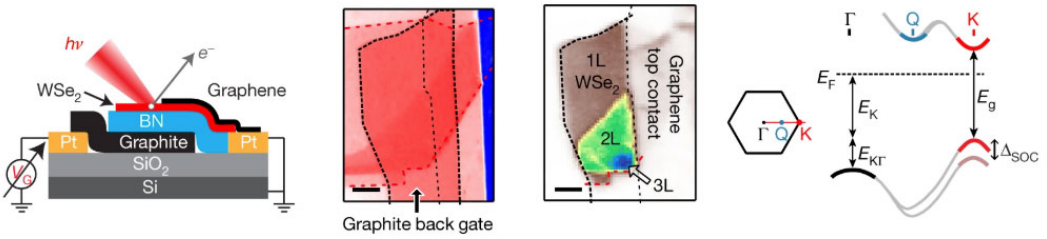
using micrometre-scale, angle-resolved photoemission spectroscopy [122–124] (microARPES) to measure band structure as a gate voltage is applied. Initially DFT was used to predict the experimental measurements, but given the high energy resolution of the experiments, it proved to be inadequate to fully explain features of the bandstructure. Therefore state of the art GW approximation, in which the theoretical considerations are detailed in chapter 4, was used instead to produce high quality band structure that is comparable with the experimental results.

The experimental procedures and results are briefly discussed at the start of this chapter, concentrating on the WSe_2 monolayer, bilayer and trilayer band structure measurements. We refer the reader to the original paper [1] for the full experimental results and discussion. After the experimental results section, the final DFT and GW results are presented and compared with the experimental measurements. In what follows an in-depth discussion of the methods used to produce the final GW band structures is presented. Finally, the chapter will end with applications to WS_2 and alloying with the virtual crystal approximation (VCA) for $Mo_{1-x}W_xS_2$.

6.2 Experimental Measurements of WSe_2

ARPES is a direct experimental technique to observe the distribution of the energy and momentum of electrons photoemitted from a solid that is illuminated by a beam of ultraviolet or X-ray light. ARPES uses a large electrostatic hemispherical electron energy analyzer that is placed on top of and parallel to the sample, along with a charge-coupled device camera, to obtain both the energy and the angle of the photoemitted electrons simultaneously [125]. Using these two information can give the energy and momentum of valence electrons in the solid sample, and hence the band structure and Fermi level. ARPES is well suited to probing 2D materials, because the electrons emitted from the excitation is only from very near the surface of the sample. Excitation spot size is typically measured in millimetres, but progress have been made in the past decade [123] to reduce the spot size with a focused beam. Micrometre-scale spot sizes, hence the term microARPES, have been achieved in at least four commissioned synchrotron beamlines by using various techniques [122–124, 126]. MicroARPES has shown promising results in studying exfoliated flakes of 2D materials that are usually tens of micrometres or fewer in size [127], and even for vertical heterostructures [128], made by stacking different 2D materials [129, 130], showing band offsets and interlayer hybridisation [131–133].

One major limitation of ARPES is that it can only probe valence electrons. In



(a) (b) (c) (d)

FIGURE 6.1: (a) Diagram of a device incorporating a WSe_2 flake, with an overlapping ground graphene top contact and gate voltage applied to the graphite back gate. (b) and (c) shows the optical and SPEM images of WSe_2 device 1 (BN thickness (d_{BN}) $= 7.4 \pm 0.5$ nm), with monolayer (1L), bilayer (2L) and trilayer (3L) regions identified. Scale bars is 5 µm. The SPEM image (b) colour corresponds to the integrated photoelectron intensity around Γ, over the full detector range of approximately 15°, corresponding to roughly 0.6 $Å^{-1}$ at 20 eV and 1.1 $Å^{-1}$ at 70 eV, and a binding-energy range of 0–3.5 eV at that point on the sample. (d) shows the Brillouin zone of MX_2 (left) and diagram of the bands along Γ–K (right), showing definitions of the energy parameters discussed in the text.

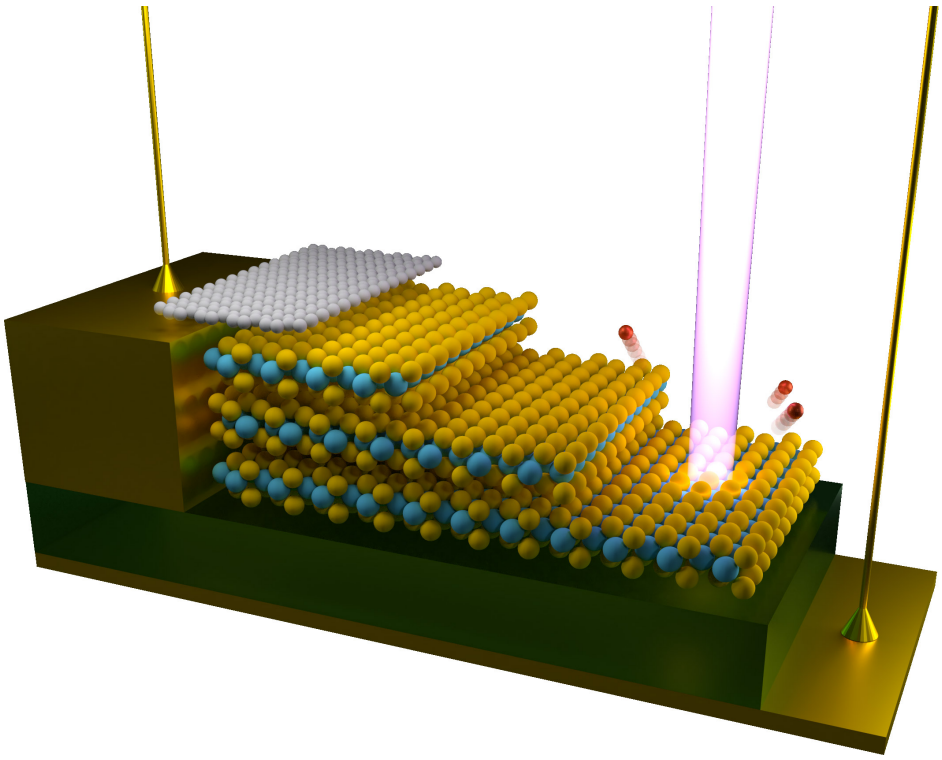


FIGURE 6.2: A 3D render of the ARPES experiments with 3 layers of WSe_2 stacked analogous to figs. 6.1a and 6.1c. The top grey atoms are a graphene layer overlaying on top of the MX_2. The two gold coloured plates and rods are Pt electrodes. The green block represents the BN layer. The red balls are electrons ejected from the WSe_2. Created as a part of this project using Blender [134].

order to measure a signal from the conduction states from a semiconductor sample, electron doping is necessary and it is usually achieved by depositing electropositive atoms such as alkali metals on the surface [129, 135–137]. This process cannot be controlled accurately and can only be reversed by high-temperature annealing; moreover, it chemically perturbs the electronic structure and introduces disorder through the random distribution of dopants. These disadvantages were removed by the use of pure electrostatic doping and thereby obtain momentum-resolved electronic spectra and direct visualisation of Fermi level shifts and band structure changes induced by applying a gate voltage. This technique was applied to 2D TMDC (MX_2) semiconductors.

It is widely believed that all monolayer MX_2 semiconductors have a direct band gap at the point K of the hexagonal Brillouin zone, but this is not known with certainty, because previous experimental measurements were unable to determine the location of the conduction-band edge (CBE) precisely. This is illustrated by the wide range of reported band gap values for monolayer WSe_2, from 1.4 eV to 2.2 eV [127, 138–142]. Furthermore, it is also uncertain when the local conduction-band minimum at the lower-symmetry point Q comes into play [138, 143].

To study gate doping of 2D semiconductors, a MX_2 flake, fabricated using standard exfoliation and polycarbonate-film-based dry transfer [144] techniques, was incorporated into the stack on top of the BN and partially overlapped by graphene, which acts as a contact to the MX_2. A gate voltage is applied to the bottom graphite layer and the top graphene layer is connected to the ground, see fig. 6.1a (a 3D rendition of this set up is shown in fig. 6.2). Figures 6.1b and 6.1c shows optical and SPEM images of a device with a WSe_2 flake that has monolayer (1L), bilayer (2L) and trilayer (3L) regions.

Due to the band gap of the MX_2, there are two cases we must consider: a small gate voltage, so that the MX_2 is insulating; and a large gate voltage, so that the MX_2 is conducting. When the gate voltage is small, the MX_2 flake is insulting enough that the BN photoconductivity brings the potential close to that of the gate. Note that although the BN is a dielectric, current can still flow through it during photoemission, due to the photoexcited carriers in the BN. The potential drop, and hence the out-of-plane electric field, across the BN will be small. In this case, there will be minimal electrons accumulating in the MX_2 flake, thus the bands will not be renormalised. For a large gate voltage, where the MX_2 is conducting, the potential drop across the BN will be close to the gate voltage, and electrons will accumulate at the CBE of the MX_2 flake as they flow in laterally from the graphene contact, and the CBE is pinned close to the graphene Fermi level.

When the flake is conducting, this constitutes a parallel-plate capacitor with geometric areal capacitance

$$C_{\mathrm{g}} = \frac{(\epsilon_0 \times \epsilon_{\mathrm{BN}})}{d_{\mathrm{BN}}} \tag{6.1}$$

where ϵ_0 is the relative permittivity of free space, $\epsilon_{\mathrm{BN}} = 4.0 \pm 0.2$ is the out-of-plane dielectric constant for BN, and d_{BN} is the thickness of the BN. During photoemission, the electrochemical potential at the emission spot will differ from ground, by an amount ΔV, associated with current flow both to the contact and to the gate which is at voltage V_{G}, thus reducing the effective gate voltage determining the local carrier density to $V_{\mathrm{G}} - \Delta V$. ΔV will not exceed the product of the effective electrical resistance, R, between the spot and ground electrode and the maximum current, which is no more than about 2 nA. Because MX_2 materials have a band gap, for small V_{G} values the in-plane resistance can be large, thus ΔV can be substantial and tracks closely to V_{G}, that is, $\Delta V \approx V_{\mathrm{G}}$, rendering the effective gate voltage close to zero. In this case, the doping concentration n_{G} is very small and the bands will not be renormalised. By contrast, R is small enough for large values of V_{G} that the electrochemical potential in the MX_2 approaches that in the ground electrode and ΔV stops changing, with the Fermi energy is effectively pinned at the CBE owing to the large density of states. Because of this, a relatively large $V_{\mathrm{G}} = 3.35\,\mathrm{V}$ was chosen for the WSe_2 measurements of the CBE. Furthermore, in this regime we can calculate the n_{G} with

$$n_{\mathrm{G}} = C_{\mathrm{g}}\left(V_{\mathrm{G}} - \frac{\Delta E_{\Gamma}}{e}\right), \tag{6.2}$$

where ΔE_{Γ} is the photoelectron kinetic energy measured relative to the Γ maximum at $V_{\mathrm{G}} = 0$ and e is the electron charge, with $V_{\mathrm{G}} - \Delta E_{\Gamma}/e$ being the static potential drop across the BN. For $V_{\mathrm{G}} = 3.35\,\mathrm{V}$, n_{G} is approximately equal to $10^{13}\,\mathrm{cm}^{-2}$. This high doping means that there is a large density of states in the flake, thus quantum capacitance [145] is negligible. Furthermore, it has been shown that quantum capacitance is not relevant in TMDCs as soon as the Fermi energy is within the conduction or valence band [146], which in our case the Fermi energy is within the conduction band. Hence, we have ignored quantum capacitance in our discussion.

The ARPES measurements were made at the Spectromicroscopy beamline of the Elettra light source [122]. Schwarzschild objective was used to focus a linear polarized light to a roughly 0.6-µm-diameter spot and it is aimed at 45° to the sample. The photon energy was 27 eV. The analyser permitted a resolution of approximately 50 meV in energy and $0.03\,\mathrm{Å}^{-1}$ in momentum. Figure 6.3 shows momentum slices obtained with the beam spot on each of these regions and along Γ–K in the WSe_2 Brillouin zone at a temperature of 100 K (fig. 6.1d, inset). As

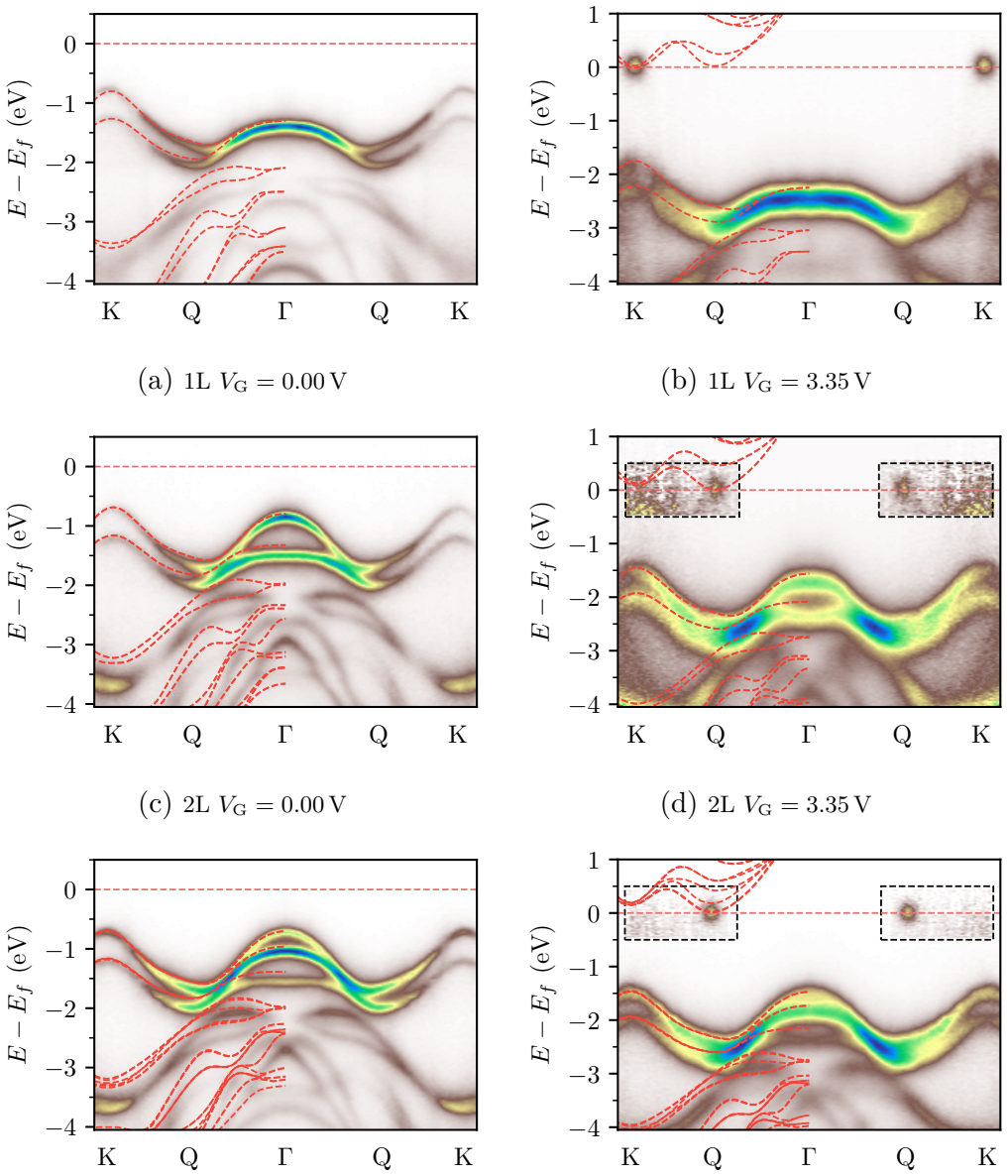
	Δ_{SOC} (eV)	E_{K} (eV)	$E_{\mathrm{K}\Gamma}$ (eV)	$m^*_{\mathrm{K}}/m_{\mathrm{e}}$	E_{g}(eV)
1L	0.485 ± 0.010	0.80 ± 0.01	0.62 ± 0.01	0.42 ± 0.05	1.79 ± 0.03
2L	0.501 ± 0.010	0.75 ± 0.01	0.14 ± 0.01	0.41 ± 0.05	1.51 ± 0.03 *
3L	0.504 ± 0.010	0.74 ± 0.01	0.00 ± 0.01	0.40 ± 0.05	1.46 ± 0.03 *

TABLE 6.1: Measured band structure parameters of WSe_2 1L, 2L and 3L. The Δ_{SOC} is the spin-orbit splitting of the valence band at K; E_{K} is the VBE at $V_{\mathrm{G}} = 0$; $E_{\mathrm{K}\Gamma} = E_{\mathrm{K}} - E_{\Gamma}$ is the difference the between the VBEs at K and Γ at $V_{\mathrm{G}} = 0$; m^*_{K} is the effective mass of the VBE at K in units of the free electron mass m_{e}; and E_{g} is the band gap measured at gate-induced electron density $n_{\mathrm{G}} = 1.0 \pm 0.2 \times 10^{12}\,\mathrm{cm}^{-2}$.
*Indirect measurement, with CBE at Q.

	$\Delta^{\mathrm{DFT}}_{\mathrm{SOC}}$ (eV)	$\Delta^{\mathrm{GW}}_{\mathrm{SOC}}$ (eV)	$E^{\mathrm{DFT}}_{\mathrm{K}\Gamma}$ (eV)	$E^{\mathrm{GW}}_{\mathrm{K}\Gamma}$ (eV)	$E^{\mathrm{DFT}}_{\mathrm{g}}$ (eV)	$E^{\mathrm{GW}}_{\mathrm{g}}$ (eV)
1L	0.464	0.478	0.50	0.56	1.27	2.31
2L	0.474	0.465	0.11	0.08	1.19	2.27 *
3L	0.489	0.481	0.02	-0.02	1.11	1.48 *

TABLE 6.2: DFT and GW band structure parameters of WSe_2 1L, 2L and 3L. The Δ_{SOC} is the spin-orbit splitting of the valence band at K; $E_{\mathrm{K}\Gamma} = E_{\mathrm{K}} - E_{\Gamma}$ is the difference the between the VBEs at K and Γ; and E_{g} is the band gap. The superscript denotes that the value is from either DFT or GW.
*Indirect gap with CBE at Q.

expected, at $V_{\mathrm{G}} = 0$ (left column) only the valence bands can be seen. Their evolution with layer number is consistent with the literature [147]. At $V_{\mathrm{G}} = 3.35\,\mathrm{V}$ (right column), an additional spot appears near E_{F}. The size of this conduction-band feature is determined solely by the resolution of the measurement. In 1L WSe_2 the spot is located at K, whereas in 2L and 3L it is at Q (see fig. 6.1d for the location of this point Q). This is consistent with evidence from photoluminescence [143] that the gap is direct at K in the monolayer, but indirect for two or more layers. Table 6.1 displays the band parameters for 1L–3L WSe_2. It shows the measured band gap, $E_{\mathrm{g}} = E_{\mathrm{C}} - E_{\mathrm{K}}$, where E_{C} is the energy of the CBE and E_{K} is the energy of the valance-band edge (VBE). It also lists the determined hole effective mass m^*_{K}, the spin–orbit splitting Δ_{SOC}, and the difference the between the VBEs at K and Γ at $V_{\mathrm{G}} = 0$ ($E_{\mathrm{K}} - E_{\Gamma}$) $E_{\mathrm{K}\Gamma}$ (the latter three being defined in fig. 6.1d)—all measured for the first time on a hexagonal BN substrate with no cap and with greater precision than in previous reports. We refer the reader to the original paper [1] for other TMDCs (MoS_2, $MoSe_2$ and WS_2) band structure parameters.

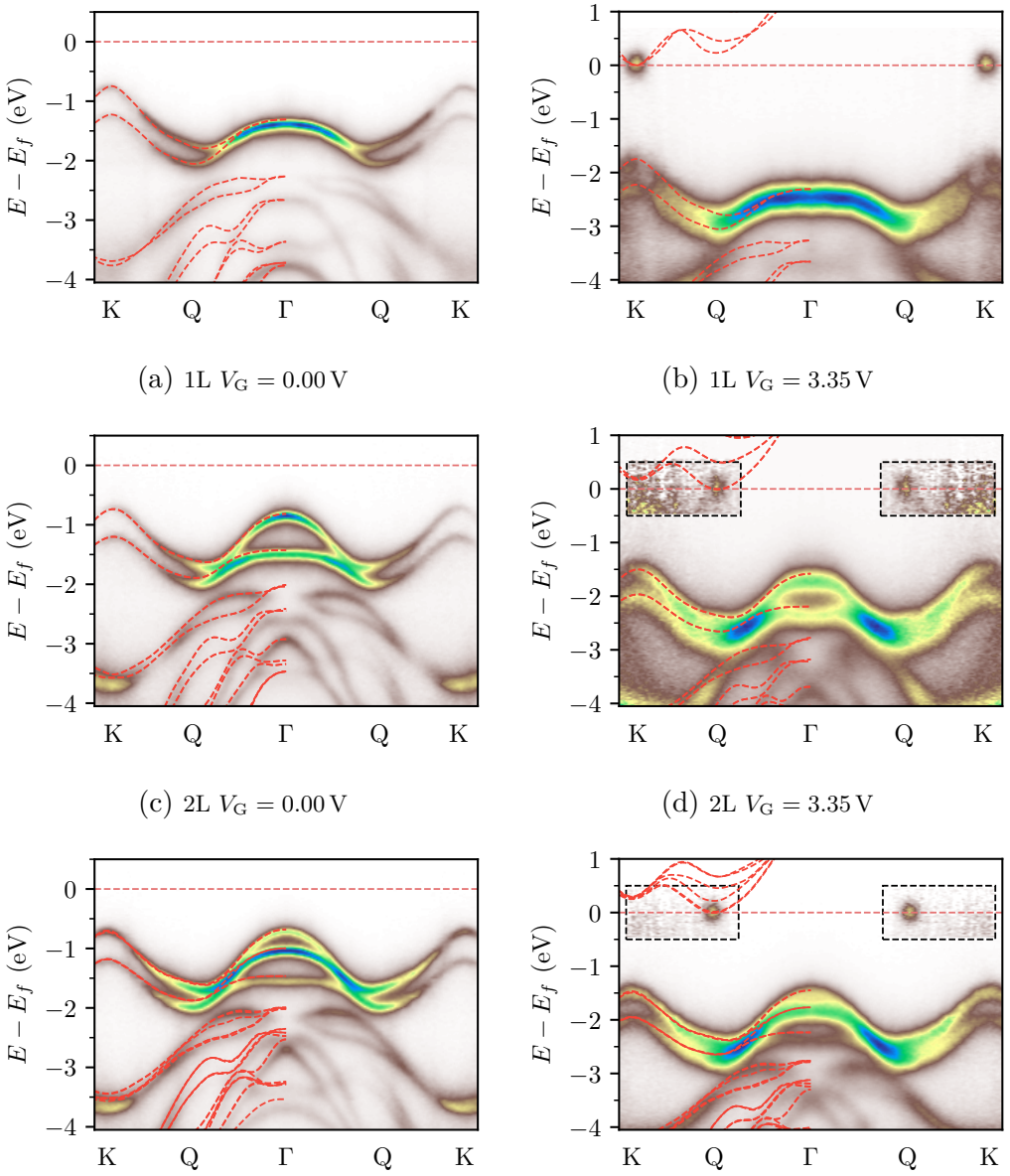


(a) 1L $V_G = 0.00\,V$ (b) 1L $V_G = 3.35\,V$

(c) 2L $V_G = 0.00\,V$ (d) 2L $V_G = 3.35\,V$

(e) 3L $V_G = 0.00\,V$ (f) 3L $V_G = 3.35\,V$

FIGURE 6.3: WSe_2 ARPES measurements along Γ-K for 1L (top row), 2L (middle row) and 3L (bottom row) regions. The left panels are at $V_G = 0$ and the right panels are at $V_G = 3.35\,V$. The intensity in the dashed box is multiplied by 20. The red dashed lines overlay are DFT calculated band structures. The conduction bands have been shifted to match the CBE of the ARPES.



(a) 1L $V_G = 0.00\,V$ (b) 1L $V_G = 3.35\,V$

(c) 2L $V_G = 0.00\,V$ (d) 2L $V_G = 3.35\,V$

(e) 3L $V_G = 0.00\,V$ (f) 3L $V_G = 3.35\,V$

FIGURE 6.4: WSe_2 ARPES measurements along Γ-K for 1L (top row), 2L (middle row) and 3L (bottom row) regions. The left panels are at $V_G = 0$ and the right panels are at $V_G = 3.35\,V$. The red dashed lines overlay are GW calculated band structures. The intensity in the dashed box is multiplied by 20. The conduction bands have been shifted to match the CBE of the ARPES.

6.3 Comparisons with the DFT and GW band structures of WSe_2

Turning our attention to the theoretical results, the DFT calculations were made using the Quantum Espresso package [44]. The optimised norm-conserving fully relativistic pseudopotentials [148] with the PBE functional [33] from PseudoDojo [149] were used. All the structures were first optimised using cell relaxation until forces were smaller than $10^{-4}\,\mathrm{Ry\,Bohr^{-1}}$. This results in a lattice constant of 3.32 Å for all three structures. Geometry optimisations and the self-consistent calculations were performed with an 18×18 in-plane k-point grid. To avoid interaction between periodic images, the vacuum spacing was 25 Å. The geometry optimisations for the 2L and 3L were performed using the optB88-vdW functional [38] to accurately describe the interlayer distance, but due to Yambo not supporting the optB88-vdW functional, the self-consistent calculations were then performed using PBE.

The red dashed lines overlay on fig. 6.3 are the results of the DFT calculations. Qualitatively, the bands have the right shapes, and it shows the 1L having a direct gap at K, while the 2L and 3L have an indirect gap where the CBE is at Q, which matches the ARPES measurements well. However, two notable problems can be seen: from the second valence bands from the VBE onwards are all shifted upwards from the ARPES, in other words, their energies are overestimated; and it is almost unclear that for 1L the band gap is in fact direct, due to the energy of the band at Q being very close to the Fermi level. We will see in the following sections that these problems are remedied by the use of the GW approximation.

Without going into the details of the parameters now, the GW calculations were made using the Yambo code [94], which used the DFT calculations as a starting point. The ARPES results with GW band structure overlaid on top are shown in fig. 6.4. Note that the conduction bands are also shifted to match the CBE of the ARPES here, as we are not expecting the GW band gap to match the experiment (more details on this later) and that we are more interested in a more direct comparison on where the CBE is. In all three structures (1L, 2L and 3L), the GW results show drastic improvements over DFT. Most notably the bands are all shifted correctly and are quantitatively comparable to experiment. The 1L conduction bands at Q also moved away from the Fermi level, thus the direct band gap nature of 1L WSe_2 is shown with certainty. Although while the 1L WSe_2 band structure maps onto the ARPES almost exactly, the 2L and 3L band structures does not match quite as well. This is attributed to the increase in the number of atoms in the system, which necessitates much tighter convergence.

Table 6.2 shows the band parameters for 1L–3L WSe_2 derived from the band structure calculations. It lists the band gap E_g, the spin–orbit splitting Δ_{SOC}, and the difference the between the VBEs at K and Γ ($E_K - E_\Gamma$) $E_{K\Gamma}$. The 1L GW results show an overall improvement over DFT. The GW spin-orbit splitting widened from the DFT and pushed the VBE at Γ down, which brought them closer to the experimental values. The band gap on the other hand is still far from the experimental value, but this is not surprising as the experimental setup has two main differences: there is band gap renormalization due to the static doping as a result of free-carrier screening [150]; and the WSe_2 is placed on a substrate which can also cause band gap renormalization [151], whereas our simulation is in a vacuum.

On the contrary, the 2L and 3L results show no improvements over DFT, but this is again could be attributed to the lack of convergence. However, it can be seen that GW seems to affect the lower energy valence bands more, leaving the highest occupied molecular orbital (HOMO) band almost unchanged, hence the minimal differences between the DFT and GW parameters. Interestingly, the 3L band gap is in excellent agreement with the experiment. This is believed to be due to the 3L stacking has removed any substrate interaction as the ARPES beam is aimed at the top of the system, thus our simulation of the system in a vacuum is now more similar to the experiment.

To bring the theoretical band gap closer to the experimental results, some interesting calculations that one can explore are the DFT and GW calculations with electron doping. We will briefly discuss some theoretical aspects of such calculations, but bare in mind that results with electron doping were not obtained for this thesis. For DFT, electron doping for the TMDCs family has been already been studied [146]. As the electron doping concentration increases, electrons will start to fill the CBE, which pushes the LUMO band down, leaving the valence bands unchanged. Hence, decreasing the band gap. However, from ref. [146], for doping concentration of $10^{13}\,\mathrm{cm}^{-2}$, it will only renormalise the band gap by a small amount ($< 0.05\,\mathrm{eV}$).

GW on the other hand, band gap renormalisation can be large due of the inclusion of screening in the exchange self-energy. As electrons increase the effects of screening also increases, which in turn affect not only the conduction bands, but also the valence bands. In GW, the valence band maximum increases with electron doping concentration, and the conduction band minimum, similar to that of DFT, decreases with electron doping concentration [150]. For monolayer MoS_2 with a doping concentration of $10^{13}\,\mathrm{cm}^{-2}$, it has been shown to reduce the band gap by $0.4\,\mathrm{eV}$ compared to without doping [150]. For monolayer WSe_2, we are likely to see the same ballpark reduction of the band gap due to its similarity with MoS_2, which

will bring the band gap much closer to the experimental results.

In this section, we have shown that GW provides quantitatively comparable band structure to the ARPES measurements, although better convergence, especially for the 2L and 3L WSe_2, is believed to be able to further improve the agreement between the GW and experimental band structure parameters. The upcoming sections will detail the procedures in producing the GW results.

6.4 Semi-Core States Within the GW Approximation

Pseudopotentials are typically used to the calculation to include only the outermost shell of an atom, which reduces the computational cost and removes problems associated with the rapid oscillatory behaviour that core electrons, as discussed in section 2.7. For DFT, reducing the number of valence electrons will generally reduce the accuracy. For GW, however, it is more complicated than just the accuracy. As it has been shown that semi-core states in certain semiconductors [152–154] and metals [155, 156] have strong influence on GW calculations, because of the strong spatial wavefunction overlap between valence and semi-core states [157]. Since exchange interactions are due to wavefunctions overlap if the exact nonlocal exchange potential is used, the exchange self-energy is therefore very sensitive to the inclusion of semi-core states. We will demonstrate in this section that the inclusion of certain semi-core states in the pseudopotential is extremely important in a GW calculation.

We have chosen to use the optimised norm-conserving fully relativistic pseudopotentials [148] from PseudoDojo [149] for all of our GW calculations, as it is well tested with the Δ-Gauge tests [149, 158]. PseudoDojo provides two sets of pseudopotentials: *standard* and *stringent*, where some elements retain more electrons as valence electrons in the stringent set. For W, there are no differences between the two sets and both contain the $5s^2$, $5p^6$, $5d^4$ and $6s^2$ states as valence electrons. For Se, both contain the $3d^{10}$, $4s^2$ and $4p^4$ states, with the stringent set also retaining the semi-core states $3s^2$ and $3p^6$ as valence electrons.

Figure 6.5a shows the band structures calculated using the standard set. As can be seen, although the DFT band structure does not show any notable problems, the GW band structure displays 'bumpy' numerical artefacts between the high symmetry BZ points, along with an indirect band gap that is not predicted by DFT. By contrast, the inclusion of the semi-core states $3s$ and $3p$ in the stringent set for Se shows none of these problems, as illustrated in fig. 6.5b. Both calculations used the same DFT and GW parameters, which are discussed in the next two sections, thus showing the impact that the semi-core states have in the GW calculation. Inclusion

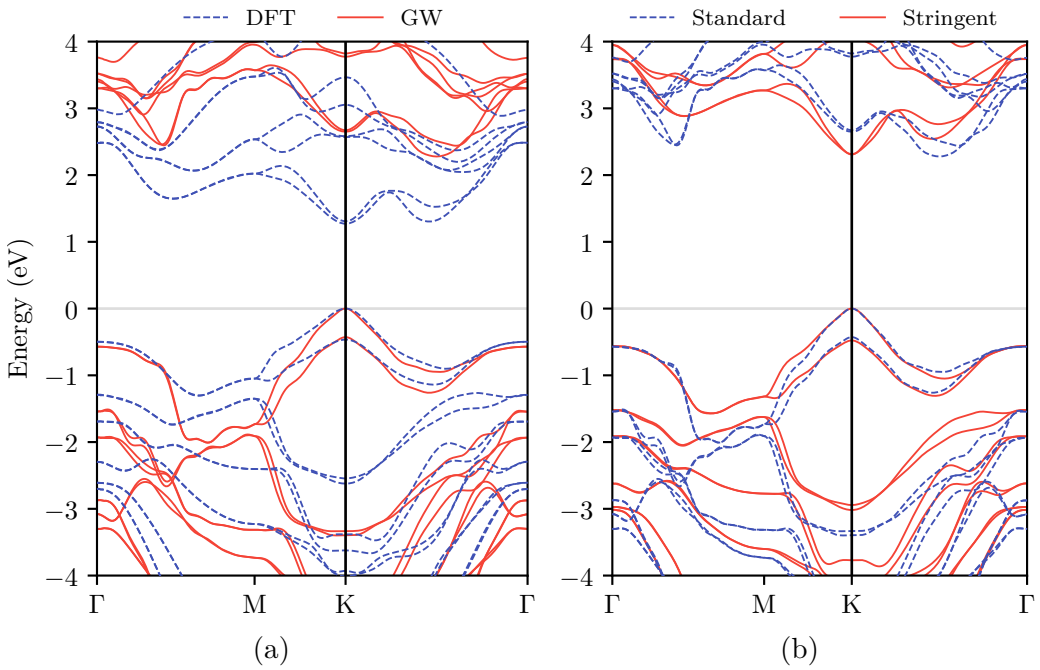


FIGURE 6.5: 1L WSe_2 band structures along Γ-M-K-Γ, where (a) compares the DFT and GW calculations using the PseudoDojo standard pseudopotentials and (b) compares the PseudoDojo standard pseudopotentials with their stringent pseudopotentials which includes the $3s$ and $3p$ semi-core states for Se.

of the $3s$ and $3p$ states is therefore concluded to be necessary to accurately compute the self-energy, and hence the stringent pseudopotentials were used to calculate all the WSe_2 GW band structures in this chapter.

6.5 The Random Integration Method

This section details the methodological concept underpinning Yambo [94] and how Yambo solves a fundamental problem of GW, the divergence of the Coulomb potential, using a novel numerical method.

The self-energy Σ from the GW approximation is often split into separate exchange (x) and a correlation (c) terms:

$$\Sigma_{n\mathbf{k}}(\omega) = \Sigma^{\mathrm{x}}_{n\mathbf{k}} + \Sigma^{\mathrm{c}}_{n\mathbf{k}}(\omega). \tag{6.3}$$

Within Yambo, the exchange term is expressed as

$$\begin{aligned}\Sigma^{\mathrm{x}}_{n\mathbf{k}} &= \langle n\mathbf{k}|\Sigma^{\mathrm{x}}|n\mathbf{k}\rangle \\ &= -\sum_m \int_{\mathrm{BZ}} \frac{d\mathbf{q}}{(2\pi)^3} \sum_{\mathbf{G}} v(\mathbf{q}+\mathbf{G})|\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G})|^2 f_{m(\mathbf{k}-\mathbf{q})},\end{aligned} \tag{6.4}$$

which is just the Fock term of the Hartree-Fock (HF) self-energy [20], where $\mathbf{G}$ are the reciprocal lattice vectors, $\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G}) = \langle n\mathbf{k}|e^{i(\mathbf{q}+\mathbf{G})\cdot\mathbf{r}}|m\mathbf{k}-\mathbf{q}\rangle$, $v(\mathbf{q}+\mathbf{G}) \equiv 4\pi/|\mathbf{q}+\mathbf{G}|^2$, and $f_{m(\mathbf{k}-\mathbf{q})}$ being the occupation factor. Note that the summation is only over occupied bands. The correlation term is given by

$$\begin{aligned}\Sigma^{\mathrm{c}}_{n\mathbf{k}}(\omega) &= \langle n\mathbf{k}|\Sigma^{\mathrm{c}}|n\mathbf{k}\rangle \\ &= i\sum_m \int_{\mathrm{BZ}} \frac{d\mathbf{q}}{(2\pi)^3} \sum_{\mathbf{G},\mathbf{G}'} v(\mathbf{q}+\mathbf{G})\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G})\rho^*_{nm}(\mathbf{k},\mathbf{q},\mathbf{G}') \\ &\quad \times \int d\omega' G^0_{m\mathbf{k}-\mathbf{q}}(\omega-\omega')\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega'),\end{aligned} \tag{6.5}$$

where, in contrast with the exchange term, the summation over bands extends over all bands, including the unoccupied ones. G^0 and $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}$ are the non-interacting Green's function and the inverse dielectric matrix as discussed in chapter 4.

The integral over the BZ in practice is replaced by a finite discretised grid of $\mathbf{q}$ points. Taking eq. (6.4) as an example, it can be rewritten as

$$\Sigma^{\mathrm{x}}_{n\mathbf{k}} \approx -\frac{\Omega_{\mathrm{BZ}}}{N_q} \sum_m \sum_{\mathbf{q}} \sum_{\mathbf{G}} v(\mathbf{q}+\mathbf{G})|\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G})|^2 f_{m(\mathbf{k}-\mathbf{q})}, \tag{6.6}$$

where Ω_{BZ} is the phase space volume of the BZ. The key assumption introduced here is that the integrand of eq. (6.4) is a smooth function of $\mathbf{q}$, which is generally true for the oscillators ρ, but which is not true for the Coulomb potential, which diverges for $\mathbf{q} \to 0$ and $\mathbf{G} \to 0$. This divergence behaviour, however, does not affect three-dimensional systems because the phase space volume associated with the Γ point reduces to $|\mathbf{q}|^2$ as $\mathbf{q} \to 0$, therefore the divergence is removed. In low-dimensional systems such as slabs and nanowires, we must still use the same 3D Coulomb potential since the systems are still three-dimensional, thus the divergence cannot be ignored, as opposed to, for example, using a 2D Coulomb potential $2\pi/\mathbf{q}$, where it is integrable in 2D. One can still remove this divergence behaviour by the use of a three-dimensional $\mathbf{q}$-point grid, but this is clearly inconvenient as it makes the calculations more costly.

To avoid the divergence problem for low-dimensional systems, Yambo has opted to use the so-called *random integration method* (RIM) [94]. Instead of evalu-

ating the Coulomb potential at each $\mathbf{q}$ point, the potential term is replaced by an integral over a new region, which we will call a small Brillouin zone (sBZ), around each $\mathbf{q}$ point. Using the HF self-energy again as an example, in the RIM it is rewritten as

$$\Sigma_{n\mathbf{k}} \approx -\sum_{m}\sum_{\mathbf{q}}\sum_{\mathbf{G}} |\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G})|^2 f_{m(\mathbf{k}-\mathbf{q})} \int_{\mathrm{sBZ}(\mathbf{q})} \frac{d\mathbf{q}'}{(2\pi)^3} v(\mathbf{q}+\mathbf{q}'+\mathbf{G}). \tag{6.7}$$

A sBZ is first constructed around the Γ point and is translated to the $\mathbf{q}$ position at each $\mathbf{q}$ point, which is represented as sBZ($\mathbf{q}$) in the above equation. The sBZ are chosen such that they cover the whole BZ. Since the integral diverges also only when $\mathbf{G} \to 0$, therefore one only need to evaluate the Coulomb integral using the RIM up to some number of $\mathbf{G}$ vectors. Hence, in the RIM, Yambo calculates the integral only for a small set of $\mathbf{G}'$ vectors:

$$v_{\mathrm{RIM}}(\mathbf{q},\mathbf{G},\mathbf{G}') = \int_{\mathrm{sBZ}(\mathbf{q})} \frac{d\mathbf{q}'}{(2\pi)^3} \frac{4\pi}{|\mathbf{q}+\mathbf{q}'+\mathbf{G}||\mathbf{q}+\mathbf{q}'+\mathbf{G}'|}. \tag{6.8}$$

In practice, the integral is computed using a three-dimensional Monte Carlo technique, where an iterative procedure which generates N random points in the sBZ is used, and the integral is evaluated as:

$$v_{\mathrm{RIM}}(\mathbf{q},\mathbf{G},\mathbf{G}') = \frac{\Omega_{\mathrm{sBZ}}}{N} \sum_{i=1}^{N} \frac{2}{|\mathbf{q}+\mathbf{q}_i+\mathbf{G}||\mathbf{q}+\mathbf{q}_i+\mathbf{G}'|} \tag{6.9}$$

Since this is just a simple summation, a large number of random points of the order of 10^6 is recommended. To converge the number of random points and $\mathbf{G}$ vectors used to evaluate the integral, one need to make sure that the RIM Coulomb integral is approximately the same as the bare Coulomb integral at the edge of the BZ. It was found that 3×10^6 random points and 100 $\mathbf{G}$ vectors is sufficient for our calculations.

To demonstrate the importance of the RIM, the HF energy gap at k-point K with and without RIM for a 1L WSe_2 for various k-grid sizes were calculated. The results are shown in fig. 6.6. Notice that with RIM the HF gap stays fairly constant (within 0.02 eV for k-grid above 12×12), while without RIM the HF gap diverges for large BZ sampling sizes. These instabilities grow as a function of the BZ sampling size because the $\mathbf{q}$ points approach the singularity.

We have also compared the GW band structure with and without RIM, as shown in fig. 6.7. The 'without RIM' valence bands are almost exactly the same as valence bands with RIM, so it displays no differences except for the band gap. The

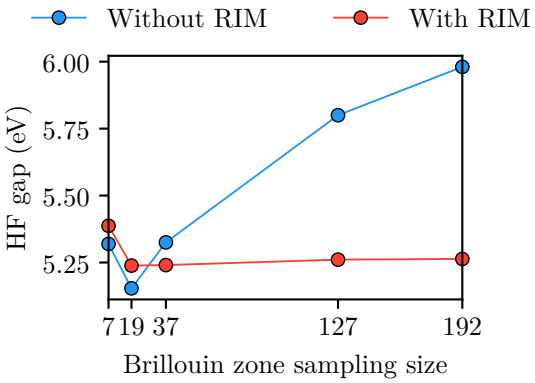


FIGURE 6.6: HF energy gap at k-point K of a 1L WSe_2 as a function of the number of k-points used to sample the BZ, with and without the RIM. The equivalent k grid sizes used here are 6×6, 12×12, 18×18, 36×36 and 45×45.

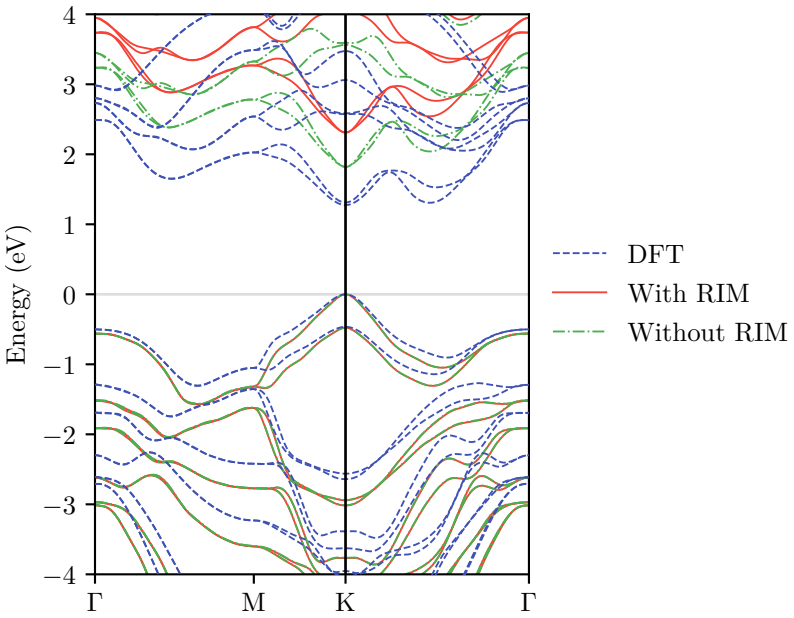


FIGURE 6.7: 1L WSe_2 band structures comparing with and without the RIM. The without RIM valence bands in green dotted dashed lines are directly on top of the with RIM valence bands in solid red lines.

band gap without RIM is in fact closer to the nominal experimental value, but due to the numerical instabilities shown in fig. 6.6, this can be regarded as a coincidence.

The RIM is used for all further GW calculations in this chapter.

6.6 Convergence

We will begin with the convergence of the exchange term eq. (6.4). The main parameter that one can change is the number of the reciprocal lattice vectors $\mathbf{G}$ it sums over. Although the $\mathbf{G}$ vectors are from inherited the DFT calculation, Yambo imposes an upper limit on the number of $\mathbf{G}$ vectors one can use. As we shall see from our convergence tests, using the full set of $\mathbf{G}$ vectors from DFT is in fact not needed, as many-body calculations in general are less sensitive to the $\mathbf{G}$ vectors than DFT ones. For our 1L WSe_2, the maximum that we can use is 16 705 vectors, or equivalently one can use an energy cutoff (see section 2.6), which in this case is 72 Ry. Since this parameter does not affect Σ^c, there is no need to compute the full GW self-energy to test for convergence. The HF energy gap at k-point K was used as the convergence test and the result is shown in fig. 6.8. As can be seen that 60 Ry is very close to convergence, within 0.01 eV difference. We have simply chosen to use the maximum $\mathbf{G}$ vectors allowed for all our calculations, because from eq. (6.4), one can see that this calculation is computationally cheap, thus using the maximum number imposed by Yambo will have minimal computational performance impact. The only downside to this is the increase in memory usage, which was not a concern for us.

The convergence of the correlation term requires two main parameters: the $\mathbf{G}$ vectors used in the screening, which determines the size of the dielectric matrix $\epsilon^{-1}_{\mathbf{G}\mathbf{G}'}$ and the number of bands, m, it sums over. For both parameters, the QP energy gap at k-point K were calculated. 300 bands was used for the convergence of the dielectric matrix size. For the number of bands, fig. 6.9b shows that the convergence is rather slow and many bands are needed to get converged results. Nevertheless, the band gap is in agreement with the reported range of 2.02–2.37 eV [99, 140, 159, 160]. Since this term is the main bottleneck of a GW calculation and we are not seeking for absolute convergence, 300 bands were used for all the 1L calculations. 300 bands were also used for the convergence of the dielectric function. Figure 6.9a shows the QP gap as we increase the cutoff energy, and it can be seen that at 5 Ry the QP gap is sufficiently converged. It should be borne in mind that if one wishes to achieve full convergence, these two parameters must be converged simultaneously (i.e., one should converge the QP gap for different dielectric plane wave energy cutoff

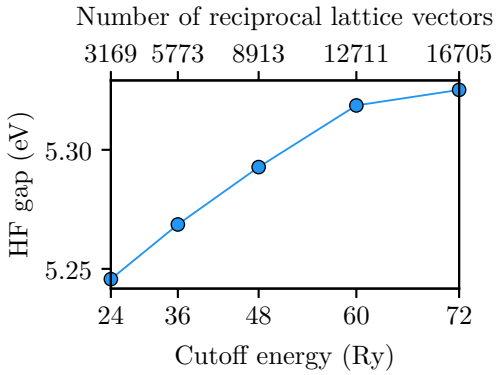


FIGURE 6.8: 1L WSe_2 HF energy gap at k-point K convergence test with the number of reciprocal lattice vector for the exchange term in the self-energy.

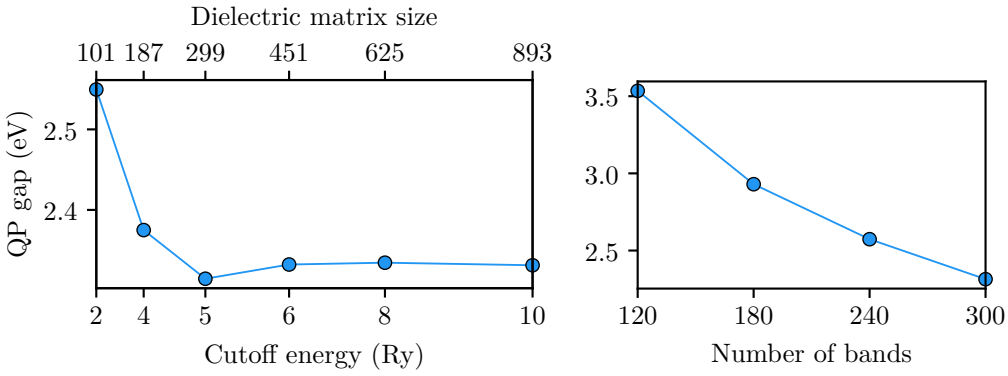
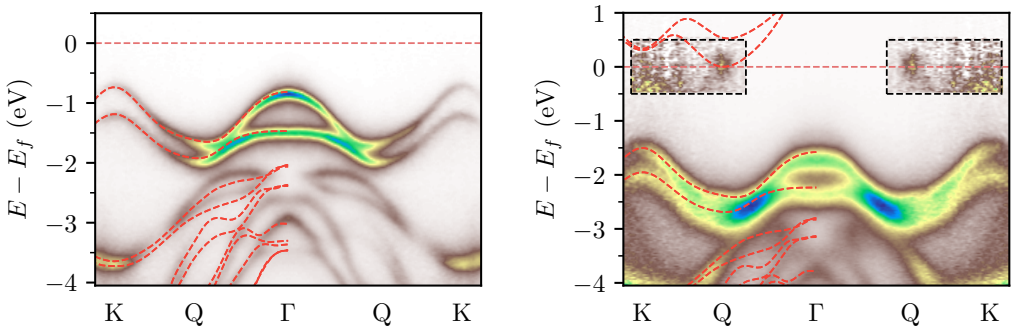


FIGURE 6.9: 1L WSe_2 GP energy gap at k-point K convergence test for the dielectric matrix size and the number of bands.



(a) $V_G = 0.00\,V$ (b) $V_G = 3.35\,V$

FIGURE 6.10: ARPES measurements along Γ-K for 2L WSe_2. The intensity in the dashed box is multiplied by 20. The red dashed lines overlay are GW band structures without EET. The conduction bands have been shifted to match the CBE of the ARPES.

while also varying the number of bands), as it has been shown [161, 162] that these two parameters are not totally independent. This concludes all the methodologies used for the 1L WSe_2 GW band structure calculation.

6.7 The Effective Energy Technique

The 2L WSe_2 used the same parameters as the 1L, with the number of bands increased to 500 and 181 101 $\mathbf{G}$ vectors were used for the exchange self-energy. The band structure is compared with the ARPES results in fig. 6.10. It already shows considerable improvement over DFT, but it was concerning to see the rapid band movement near Γ for the third and fourth band from the Fermi level, and that the fourth band almost hybridises with the fifth band. Of course, we can increase the number of bands to achieve better convergence, but it comes at a much higher computational cost. In this section, we explore the so-called effective energy technique (EET) [84] that allows better convergence without the need to include more bands.

Firstly, the integral over BZ in the correlation term in eq. (6.5) can be replaced with a summation over $\mathbf{q}$ points. Then as we are interested in reducing the number of bands, we will separate out the terms that are summations over $(\mathbf{q}, \mathbf{G}, \mathbf{G}')$ and the terms that are summations over bands m. With the frequency integral also

replaced with a summation, which will be denoted as g, we can write

$$\Sigma^{\mathrm{c}}_{n\mathbf{k}}(\omega) = \sum_{g}^{\infty} \sum_{\mathbf{q},\mathbf{G},\mathbf{G}'} F(\mathbf{q},\mathbf{G},\mathbf{G}',\omega_g) S^{g}_{n\mathbf{k}}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega). \tag{6.10}$$

$F(\mathbf{q},\mathbf{G},\mathbf{G}',\omega_g)$ is just all the terms collected up that are not dependent on m, for example the bare coulomb potential and the inverse dielectric matrix. As these are not important to this derivation, they will be left as they are. $S^{n}_{g}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega)$ contains the summation over all bands, and with the GF written out in full, it is defined as

$$S^{g}_{n\mathbf{k}}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega) = \sum_{m} \frac{\rho_{nm}(\mathbf{k},\mathbf{q},\mathbf{G})\rho^{*}_{nm}(\mathbf{k},\mathbf{q},\mathbf{G}')}{\omega - \omega_g - \epsilon_m + i\delta}. \tag{6.11}$$

The summation over m can be split into a summation over occupied states v and a summation over unoccupied states c. Because we wish to eliminate the summation over unoccupied states, we will only consider the summation over the in principle infinite unoccupied states. By introducing a function $\Delta_{ng}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega)$, the above equation can be rewritten as

$$S^{g}_{n\mathbf{k}}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega) = \frac{\sum_c \rho_{nc}(\mathbf{k},\mathbf{q},\mathbf{G})\rho^{*}_{nc}(\mathbf{k},\mathbf{q},\mathbf{G}')}{\omega - \omega_g - \epsilon_n - \Delta_{ng}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega) + i\delta}. \tag{6.12}$$

Such a function, $\Delta_{ng}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega)$, can always be found because it has the same degrees of freedom as the left-hand side of this equation. By introducing such a function, the closure relation, $\sum_c |c\rangle\langle c| = 1 - \sum_v |v\rangle\langle v|$, can be employed, thus an expression for $S^{g}_{n\mathbf{k}}$ which contains only a summation over occupied states is obtained. It is useful to see that eq. (6.12) can be split into two parts

$$\begin{aligned} S^{g}_{n\mathbf{k}}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega) = &\sum_{c=N_v+1}^{M} \frac{\rho_{nc}(\mathbf{k},\mathbf{q},\mathbf{G})\rho^{*}_{nc}(\mathbf{k},\mathbf{q},\mathbf{G}')}{\omega - \omega_g - \epsilon_c + i\delta} \\ &+ \sum_{c=M+1}^{\infty} \frac{\rho_{nc}(\mathbf{k},\mathbf{q},\mathbf{G})\rho^{*}_{nc}(\mathbf{k},\mathbf{q},\mathbf{G}')}{\omega - \omega_g - \epsilon_n - \Delta_{ng}(\mathbf{q},\mathbf{G},\mathbf{G}',\omega) + i\delta}, \end{aligned} \tag{6.13}$$

where N_v is the number of occupied states. This allows the approximation of Δ_{ng} to be smaller as the number of unoccupied states we include, M, increases, so that we do not need to rely solely on the approximation and better convergence can still be achieved with increasing number of states. This method is parameter free and the original literature [84] supplies several approximation schemes for Δ_{ng}.

Figure 6.11 shows the comparison between calculations with and without EET band structure. The EET improved the Δ_{SOC} at K from 0.450 eV to 0.465 eV

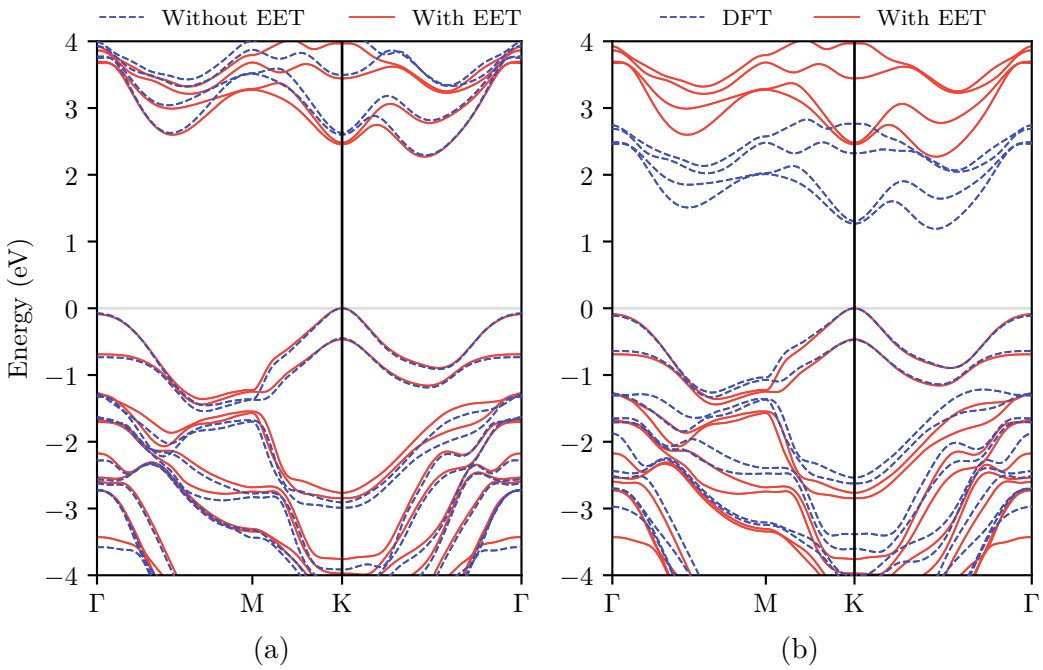


FIGURE 6.11: 2L WSe_2 band structures along Γ-M-K-Γ comparing (a) with and without the EET and (b) DFT and GW with the EET.

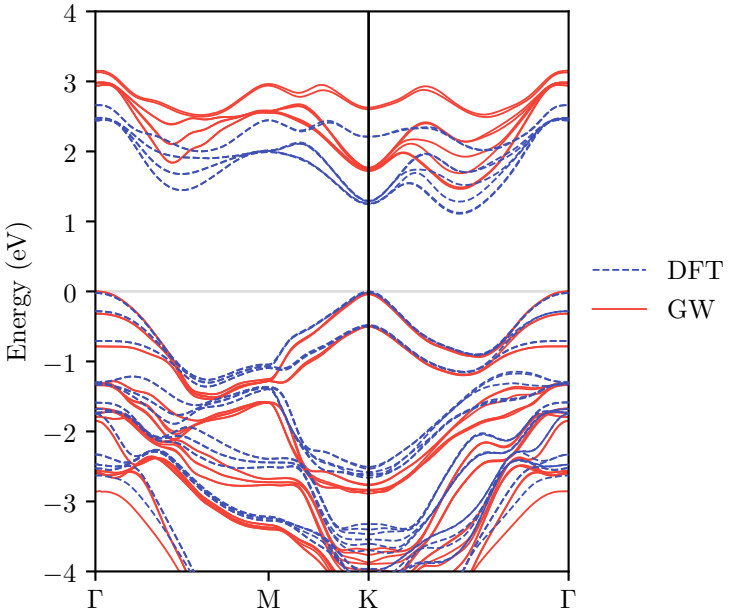


FIGURE 6.12: 3L WSe_2 band structure along Γ-M-K-Γ comparing DFT and GW.

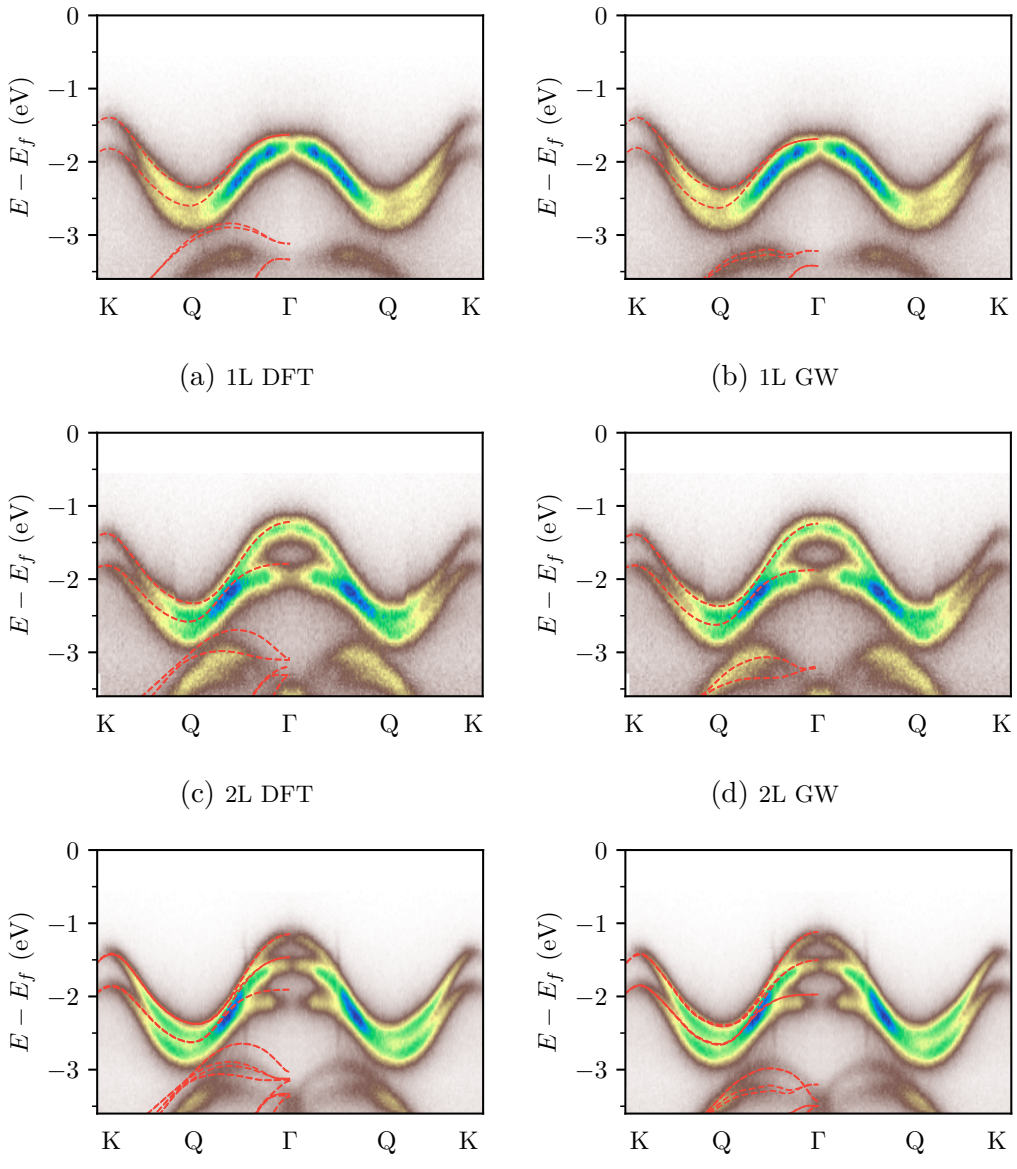
and the band gap from 2.29 eV to 2.27 eV, which are both closer to the experimental measurements. Furthermore, it improved the aforementioned problems with the bands near Γ. The EET was also employed to calculate the 3L WSe_2 band structure. Figure 6.12 shows the 3L WSe_2 band structure comparison with DFT using 138 349 **G** vectors and 700 bands. It can be seen that the VBE at Γ has been shifted up above the VBE at K, causing the band gap to be between Γ and Q. Although this technique have shown promising results without the need to include more bands into the calculation, better convergence is still needed for the 2L and 3L WSe_2.

This concludes all the parameters methodologies used to produce the final results for the 1L, 2L and 3L WSe_2 aheadly shown in fig. 6.4.

6.8 Application to WS_2

The methodologies described in all the previous sections in this chapter were also applied to WS_2. The optimised norm-conserving fully relativistic pseudopotentials [148] with the PBE functional [33] from PseudoDojo [149] were used, where W contains the $5s^2$, $5p^6$, $5d^4$ and $6s^2$ valence states, S contains the $3s^2$ and $3p^4$ valence states. All the structures were first optimised using cell relaxation until forces were smaller than $10^{-4}\,\mathrm{Ry\,Bohr^{-1}}$. The optB88-vdW functional [38] was used for the 2L and 3L. This results in a lattice constant of 3.19 Å for all three structures. An 18×18 in-plane k-point grid was used for all the DFT calculations. We used 300 bands, 500 bands and 700 bands for the monolayer, bilayer and trilayer WS_2, respectively, for the self-energy and dynamical dielectric screening. The RIM with 3×10^6 random q-points was used, along the EET for all three structures.

The ARPES measurements were performed similarly to section 6.2, but without any static doping so the CBE is not visible. The experimental results are the work of Natalie Teutsch from the University of Warwick. Figure 6.13 shows the ARPES results along with theoretical band structure overlaid on top. Unfortunately, only a small energy range was measured hence the small number of bands. In both DFT and GW the HOMO bands at Γ have slightly too high energy, similar to the WSe_2 results, with GW providing only limited improvement. In all three structures, GW has caused minimal changes to the shape of the highest energy valence bands, similarly to WSe_2. Nevertheless, the GW calculations show clear improvements over DFT.



(a) 1L DFT (b) 1L GW

(c) 2L DFT (d) 2L GW

(e) 3L DFT (f) 3L GW

FIGURE 6.13: WS_2 ARPES measurements along Γ-K for 1L (top row), 2L (middle row) and 3L (bottom row) regions. The red dashed lines overlay are the DFT (left panels) and GW (right panels) calculated band structures..

6.9 Application to $Mo_{1-x}W_xS_2$

Band gap engineering is essential to the optimisation of electronics and optoelectronics. One method to tune the band gap is alloying, where different metal elements are mixed into a single material. For TMDCs, substituting various transition metal (Mo, W, Ti, etc) and chalcogen (S, Se, Te) atoms into the TMDC lattice while maintaining the original stoichiometry would allow the modulation of certain electronic properties. Monolayer $Mo_{1-x}W_xS_2$, in which the Mo are mixed with W, have been shown to produce different band gaps ranging from ∼1.89 eV to ∼2.03 eV with increasing W concentration [163, 164]. In this section, we will compare DFT and GW band gaps of MoS_2, $Mo_{0.5}W_{0.5}S_2$ and WS_2.

The calculations are prepared in the same way as the last section. For $Mo_{0.5}W_{0.5}S_2$, the virtual crystal approximation (VCA) [165] was used to produce the pseudopotentials. In the VCA, the effect of the different potentials is included in an average potential weighted by their concentration. For example, for elements Mo and W with x being the concentration of W, the VCA potential is given by $V_{\mathrm{VCA}}(\mathbf{r}) = (1-x)V_{\mathrm{Mo}}(\mathbf{r}) + xV_{\mathrm{W}}(\mathbf{r})$. Note that although VCA has been recognised as an acceptable approximation for alloys, the results may not be accurate in some cases, as the mixing is only in the potentials and not also, for example, the Coulomb nuclear potential.

The geometry optimised lattice constants for MoS_2, $Mo_{0.5}W_{0.5}S_2$ and WS_2 are 3.185 , 3.184 and 3.186 The DFT and GW band structures are illustrated in fig. 6.14. The $Mo_{0.5}W_{0.5}S_2$ band structure shows promising results where the band gap and spin-orbit splitting Δ_{SOC} is in between MoS_2 and WS_2. Figure 6.15 highlights the Δ_{SOC} at K and the band gap. DFT predicts that the band gap decreases as W concentration increases, which is not what has been found from experiments [163, 164]. GW shows a clear improvement over DFT, where the band gap increases with W concentration. However, the literature results also note that the band gaps exhibit the so-called 'bowing effect', where it initially decreases, then gradually increases, so at 0.5 concentration the band gap increase is expected to be very small from 0 concentration (∼0.01 eV), whereas our GW gap has a more significant jump of 0.04 eV. It can also be seen that our gaps range from 2.69 eV to 2.75 eV, which is smaller than the reported range.

It has been shown that the bowing effect is attributed to the difference in the major contribution of metal elements in the LUMO between MoS_2 and WS_2 [166]. The d_{z^2} orbital is the major contribution of metal in MoS_2, whereas the d_{xy}, $d_{x^2-y^2}$ and d_{z^2} orbitals are the major contribution in WS_2. Since the metal elements

contribution to the HOMO is the same for MoS_2 and WS_2, where both are d_{xy} and $d_{x^2-y^2}$ orbitals, increasing W concentration would lead to a linear shift in the HOMO energy level, thus it does not contribute to the bowing effect. In contrast, due to the difference in the orbital composition in the LUMO and the higher LUMO energy of WS_2, W atoms contribute less to the LUMO, so the shift caused by W is non-linear. That is, when the W concentration is low, the LUMO energy level shift caused by W is minimal, and only when the W concentration is sufficiently high that the W contribution to the LUMO is significant, creating a parabolic shift in the LUMO energy level as W concentration increases. Hence, the bowing effect.

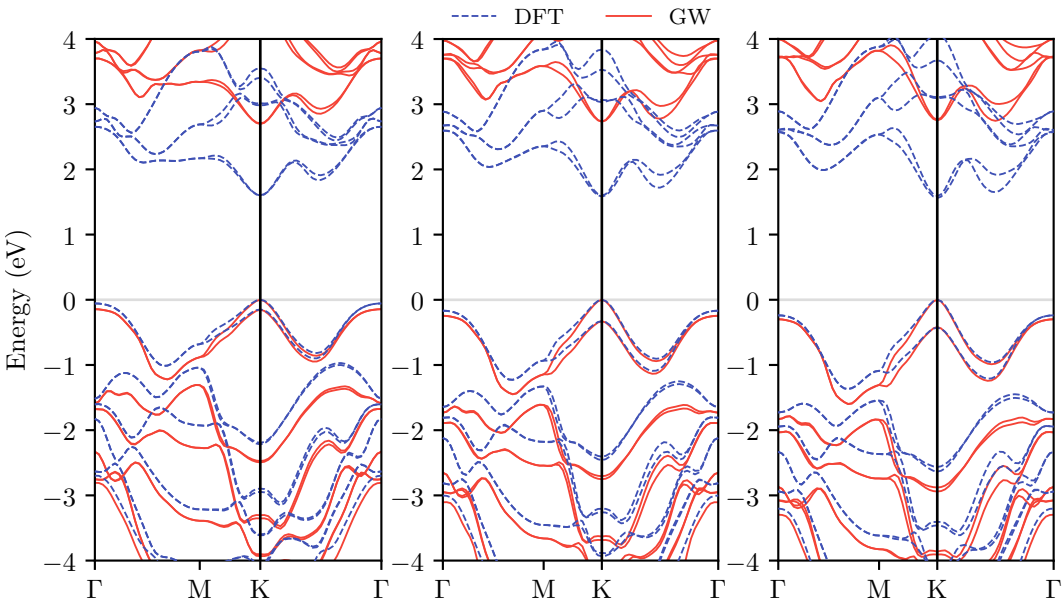
The Δ_{SOC} values show almost no difference to DFT, with WS_2 Δ_{SOC} decreasing from 0.427 eV to 0.417 eV, which puts it further from the experimental value of 0.45 ± 0.03 eV [1]. A wider range of calculations on different concentration, more rigorous convergence tests and careful analysis of the accuracy of VCA would be needed before it could be used with confidence.

6.10 Conclusions

We have presented the comparison of DFT and GW band structures with experimental ARPES measurements. In both WSe_2 and WS_2 1L, 2L and 3L, the GW band structures are in excellent agreement with ARPES, while DFT is only qualitatively comparable. Several shortcomings of GW were discovered and addressed. From our investigation, the lack of certain semi-core states in GW could cause bumpy numerical artefacts in the band structure. It was found that it is necessary to include the 3s and 3p states in Se, which are not always included, to accurately capture the self energy. We have also demonstrated that the HF gap of 2D materials diverges for large k-point sampling size if the divergence of the Coulomb term is not properly dealt with. Yambo addresses this problem by the random integration method, where random points are sampled around the singularity to produce an approximate, but accurate, integral of the bare Coulomb term.

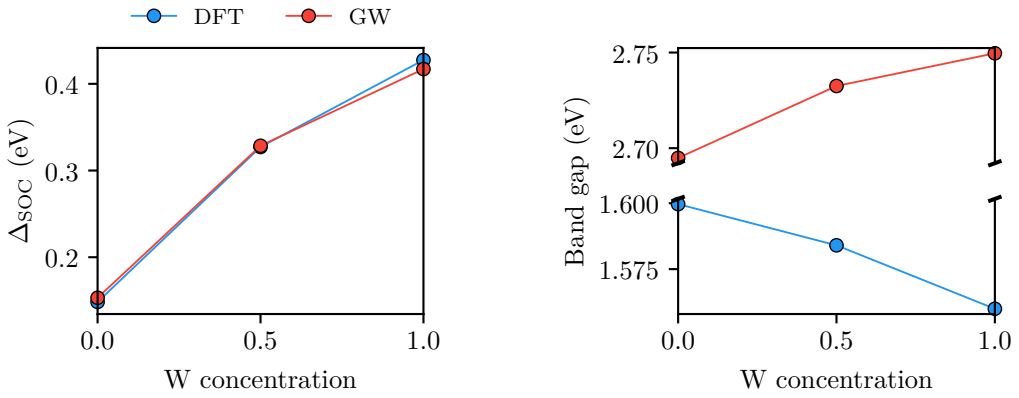
It was also observed that GW calculations converges very slowly with the number of bands. Therefore, we have tested the effective energy technique and have shown to improve convergence even without adding additional bands. Even with the refinement in the convergence, we noted that the bilayer and the trilayer results are not converged and they can certainly be improved to better match with the experimental results.

Finally, as a side project, we explored the use of the VCA to simulate the $Mo_{1-x}W_xS_2$ alloy. Initial findings have shown promising results. We have found



(a) MoS_2 (b) $Mo_{0.5}W_{0.5}S$ (c) WS_2

FIGURE 6.14: 1L band structure along Γ-M-K-Γ comparing DFT and GW for (a) MoS_2, (b) $Mo_{0.5}W_{0.5}S$ and (c) WS_2, showing band structure changes as W concentration increases in $Mo_{1-x}W_xS_2$ alloy.



(a) (b)

FIGURE 6.15: The measured band parameters for DFT and GW for $Mo_{1-x}W_xS_2$ alloy, where (a) shows the spin-orbit splitting at K and (b) shows the band gap at K.

that DFT was unable to predict the band gap increase as W concentrate increases. Conversely, GW results show that the band gap increases as W, which is expected. However, it was found that GW was ineffective at correcting the spin-orbit splitting. This work could be continued in the future, running more concentrations and make sure the calculations are converged.

This chapter has provided a recipe for running GW calculations. In summary, below are our recommendations when running GW calculations:

- If one observe 'bumpy' numerical artefacts in the band structure, one should check whether the pseudopotentials used include enough semi-core states.
- For low-dimensional materials, one could use the random integration method to deal with the divergence problem of the Coulomb potential.
- Since the calculation for the exchange self-energy is not expensive, one can simply use the maximum number of **G** vectors allowed (i.e., the same energy cutoff as the plane wave energy cutoff from the DFT starting point calculation).
- For the convergence of the correlation self-energy, the dielectric plane wave energy cutoff and the number of bands needs to be converged simultaneously, as they are not totally independent parameters.
- The effective energy technique could be used to reduce the number of bands needed to achieve convergence.

A list of all the important parameters used in this chapter is shown in Table 6.3.

Parameter	Value
Quantum Espresso	
Plane-wave cutoff energy	140 Ry
XC functional	PBE
XC functional for 2L & 3L geometry optimisation	optB88-vdW
Force tolerance for geometry optimisation	10^{-4} Ry Bohr^{-1}
Convergence threshold for estimated energy error	10^{-8} Ry
Number of bands for 1L	300
Number of bands for 2L	500
Number of bands for 3L	700
Yambo	
Number of q-points for RIM	3×10^6
G vectors for RIM	100
G vectors for exchange self-energy	Maximum available
G vectors for dielectric matrix	5 Ry
Terminator to accelerate convergence	BRS (Use the EET)
Pseudopotentials	Valence states
Mo	4s 4p 4d 5s
W	5s 5p 5d 6s
S	3s 3p
Se	3s 3p 3d 4s 4p

TABLE 6.3: A table showing several of the most important parameters used in Quantum Espresso and Yambo, and a list of pseudopotentials used which would be necessary to reproduce the calculations performed in this chapter.

Chapter 7

Species Dependent Scissor Operator and MoS_2/TiS_2 Vertical Heterostructure

7.1 Introduction

Vertical heterostructures (VHSs) consists of stacking of different layered materials, held together by van der Waals (vdW) forces. Because of the weak vdW interlayer interactions, construction of VHSs with desired properties is permitted by simply selecting the composition. VHSs have attracted much attention in both the theoretical and experimental community. They have shown to have untapped potential in electronics and optoelectronics such as field-effect transistor [167, 168] and photovoltaic solar technology [169–171].

The previous chapter have demonstrated the power of the GW approximation. GW calculations, however, are very computationally demanding, making it infeasible for calculations with more than a few atoms. On the other hand, DFT exhibits the well-known band gap problem, where DFT calculations significantly underestimate the fundamental band gap, and can lead to incorrect predictions of the band alignment of semiconductors, which is crucial for VHSs. Figure 7.1a illustrates a set of band alignments of transition metal disulfide (MS_2) predicted by DFT and GW. As can be seen, there are pairs of MS_2 that shows a type III band alignment (broken gap) from DFT, but GW shows a type II band alignment (staggered gap). For example, HfS_2/WS_2 and TiS_2/MoS_2.

The answer to whether one can use GW to study VHSs is, it depends. For materials with similar lattice constant, one can certainly use GW with a small

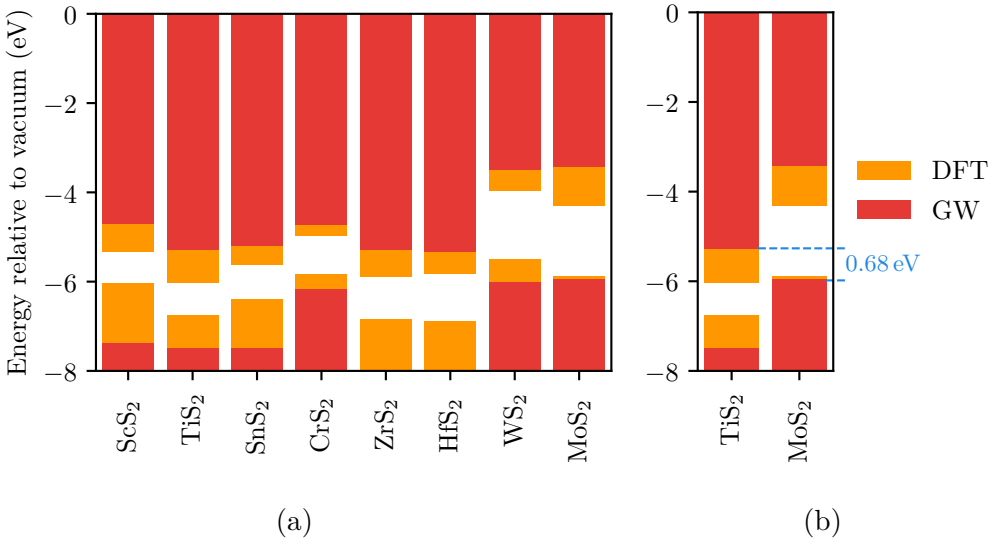


FIGURE 7.1: (a) Small set of 2H transition metal disulfide band alignments predicted by DFT (orange) and GW (red). (b) is the band alignment of TiS_2 and MoS_2 with DFT showing a broken gap but GW showing staggered band gap of size 0.68 eV. The DFT and GW data are obtained from S. Haastrup et al [159].

simulation cell with minimal strain containing only a few atoms. For materials with dissimilar lattice constant, which often produces more interesting electronic properties [131], they produce Moiré patterns [172, 173] that can only be simulated through very large supercells, thus GW would be impracticable.

This chapter explore the possibility of correcting the band alignment of VHSs using a novel scissor operator approach. We will first describe the scissor operator and its implementation to ONETEP, then the application to the TiS_2/MoS_2 VHS.

7.2 Species Dependent Scissor Operator and NGWF Gradient

The traditional scissor operator can be used to open band gap by a rigid shift of the conduction band upwards with respect to the valence band, and hence alleviates the band gap problem. Due to theoretical spectroscopy requiring quasiparticle energies which DFT lacks, the scissor operator can be used to aid DFT theoretical spectroscopy calculations, and it has shown to be useful in the calculations of optical absorption spectra [174].

Instead of just applying the shift to the conduction band, our novel scissor operator allows one to apply species-dependent and subspace-dependent energy-level

shifts to the hamiltonian, which has the effect of shifting eigenvalues associated with specific layers of the material. One can separately shift the valence and conduction subspaces associated with each layer. This also affects the total energy, so it must be applied with great care if you are using the total energy for any purpose. The idea is that a band alignment correction can be applied to the individual layers of non-covalently-bonded layered materials, though there may well be many other applications as well. It would seem 'unwise' at best to apply this approach to different regions of the same molecule or solid which are strongly bonded: results would be unpredictable and likely unphysical. One ideal use would be to correct the alignment of the band-edges of a layered material heterobilayer so that the appropriate heterostructure type was realised, for example straddled-gap rather than broken-gap, using shifts chosen by reference to beyond-DFT accuracy calculations of the individual materials, or from experimental techniques such as ARPES.

The traditional scissor operator within KS DFT can be written as an extra potential added to the Hamiltonian:

$$\hat{H}_{\text{scissor}} = \sum_{i}^{\text{val}} \sigma_{\text{v}} \left|\psi_i\right\rangle\!\left\langle\psi_i\right| + \sum_{i}^{\text{cond}} \sigma_{\text{c}} \left|\psi_i\right\rangle\!\left\langle\psi_i\right| . \tag{7.1}$$

The σ_v and σ_c are the shifts for valence and conduction states, respectively. We can use the descriptions given in chapter 3 to transform this to the NGWF formalism, which gives

$$\hat{H}_{\text{scissor}} = \sigma_{\text{v}} \left|\phi_\eta\right\rangle K^{\eta\delta} \left\langle\phi_\delta\right| + \sigma_{\text{c}} \left|\phi_\eta\right\rangle \left(S^{\eta\delta} - K^{\eta\delta}\right) \left\langle\phi_\delta\right| . \tag{7.2}$$

$S^{\eta\delta}$ is the inverse overlap matrix, so that $S^{\eta\delta} - K^{\eta\delta}$ is what we can call the conduction density kernel. Note that in the above equations and in what follows, unless otherwise stated, the Einstein notation is adopted, in which implicit summation over repeated Greek indices is assumed.

As we are interested in applying the shifts to different atoms in the system, we defined our scissor shifts to be dependent on the atom:

$$\hat{H}_{\text{scissor}} = \left|\phi_\eta\right\rangle K^{\eta\delta}_{\text{shifted}} \left\langle\phi_\delta\right| , \tag{7.3}$$

where K_{shifted} is the sum of the species-dependent shifted valence and conduction density kernel, which is defined as

$$K^{\eta\delta}_{\text{shifted}} = \sum_{L} \left(\sigma_v^L K_L^{\eta\delta} + \sigma_c^L (S_L^{\eta\delta} - K_L^{\eta\delta})\right). \tag{7.4}$$

The σ_v and σ_c are the shifts for valence and conduction states of layer L. Although in general the sum is over all atoms, so that different shifts can in fact be applied to different atoms, we use L here to reinforce the importance that this operator should only be used on layered materials where the interaction between the layers is weak.

Within the NGWF formalism, the total energy can be calculated using [61, 73]

$$E = \frac{n}{K^{\kappa\lambda}S_{\lambda\kappa}} K^{\alpha\beta} \langle \phi_\beta | \hat{H}_\mathrm{o} + \hat{H}_\mathrm{scissor} | \phi_\alpha \rangle + E_\mathrm{dc}. \tag{7.5}$$

The $\frac{n}{K^{\kappa\lambda}S_{\lambda\kappa}}$ is a scaling factor to ensure that at all times the constraint of constant number of electrons n is considered, because before the density kernel converges, $\mathrm{tr}[\mathbf{KS}] \neq n$. The Hamiltonian has been separated into two parts to illustrate what has been added: the original Hamiltonian without the scissor operator $\hat{H}_\mathrm{o}$ and the scissor operator $\hat{H}_\mathrm{scissor}$ from eq. (7.3). The last term, E_dc, is the 'double-counting' [54, 61] correction which contains the XC energy and terms compensating the double counting of the Coulomb electronic repulsion and the XC potential.

Because of the additional scissor operator in the energy term, the NGWF optimisation which optimises the expansion coefficients $c_{i,\alpha}$ of eq. (3.22) will also need to be modified. The expansion coefficients are found by minimising the energy gradient, $\frac{\partial E}{\partial \phi^*_\alpha(\mathbf{r})}$ [61, 73]. Here, each term of the energy gradient will be derived and discussed to see what needs to be added explicitly for our scissor operator.

We will begin by defining $H_{\beta\alpha} = \langle \phi_\beta | \hat{H}_\mathrm{o} + \hat{H}_\mathrm{scissor} | \phi_\alpha \rangle$, so that the energy derivative is

$$\begin{aligned}
\frac{\partial E}{\partial \phi^*_\alpha(\mathbf{r})} &= \frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})} \left[\frac{n K^{\alpha\beta}(\{\phi_\alpha\}) H_{\beta\alpha}(K^{\alpha\beta}_n, \{\phi_\alpha\})}{K^{\kappa\lambda}S_{\lambda\kappa}} \right] \\
&= \frac{\partial K^{\alpha\beta}}{\partial \phi^*_\alpha(\mathbf{r})} \frac{n H_{\beta\alpha}}{K^{\kappa\lambda}S_{\lambda\kappa}} + \frac{n K^{\alpha\beta}}{K^{\kappa\lambda}S_{\lambda\kappa}} \frac{\partial H_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \\
&\quad - \frac{n K^{\alpha\beta} H_{\beta\alpha}}{\left(K^{\kappa\lambda}S_{\lambda\kappa}\right)^2} \frac{\partial K^{\mu\nu}}{\partial \phi^*_\alpha(\mathbf{r})} S_{\nu\mu} - \frac{n K^{\alpha\beta} H_{\beta\alpha}}{\left(K^{\kappa\lambda}S_{\lambda\kappa}\right)^2} K^{\mu\nu} \frac{\partial S_{\nu\mu}}{\partial \phi^*_\alpha(\mathbf{r})} \\
&= \frac{n}{K^{\kappa\lambda}S_{\lambda\kappa}} \left(\frac{\partial K^{\alpha\beta}}{\partial \phi^*_\alpha(\mathbf{r})} (H - \mu S)_{\beta\alpha} - \mu K^{\mu\nu} \frac{\partial S_{\nu\mu}}{\partial \phi^*_\alpha(\mathbf{r})} \right) + K^{\alpha\beta}_n \frac{\partial H_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})},
\end{aligned} \tag{7.6}$$

where μ is a chemical-potential-like term given by

$$\mu = \frac{K^{\alpha\beta} H_{\beta\alpha}}{K^{\kappa\lambda}S_{\lambda\kappa}}, \tag{7.7}$$

and $K_n^{\alpha\beta}$ is the rescaled density kernel for n electrons:

$$K_n^{\alpha\beta} = \frac{nK^{\alpha\beta}}{K^{\kappa\lambda}S_{\lambda\kappa}}. \tag{7.8}$$

To differentiate $\frac{\partial K^{\alpha\beta}}{\partial \phi^*_\alpha(\mathbf{r})}$, we need to write out $K^{\alpha\beta}$ in terms of the auxiliary matrix $\mathbf{L}$ and the overlap matrix $\mathbf{S}$ using the purification transformation (eq. (3.17)) such that $K^{\alpha\beta} = (3LSL - 2LSLSL)^{\alpha\beta}$. We will also simplify the equation by defining $H' = H - \mu S$. The first term in the bracket of eq. (7.6) now reads

$$\begin{aligned}
\frac{\partial K^{\alpha\beta}}{\partial \phi^*_\alpha(\mathbf{r})}(H - \mu S) &= \frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})}(3LSL - 2LSLSL)^{\alpha\beta}H'_{\beta\alpha} \\
&= \left(3L\frac{\partial S}{\partial \phi^*_\alpha(\mathbf{r})}L - 2L\frac{\partial S}{\partial \phi^*_\alpha(\mathbf{r})}LSL - 2LSL\frac{\partial S}{\partial \phi^*_\alpha(\mathbf{r})}L\right)^{\alpha\beta}H'_{\beta\alpha} \\
&= (3LH'L - 2LH'LSL - 2LSLH'L)^{\alpha\beta}\frac{\partial S_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})}.
\end{aligned} \tag{7.9}$$

We note that the terms that are not $K^{\alpha\beta}$, in this case H', are just 'swapped' with S. This is permitted because of the use of the tensor indices. This property of the derivative is quite important and will be exploited again later.

The energy gradient in eq. (7.6) can now be written as

$$\frac{\partial E}{\partial \phi^*_\alpha(\mathbf{r})} = Q^{\alpha\beta}\frac{\partial S_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} + K_n^{\alpha\beta}\frac{\partial H_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})}, \tag{7.10}$$

with

$$Q^{\alpha\beta} = \frac{n}{K^{\kappa\lambda}S_{\lambda\kappa}}(3LH'L - 2LSLH'L - 2LH'LSL - \mu K)^{\alpha\beta}. \tag{7.11}$$

The H' and μ terms inside Q already contain the scissor Hamiltonian when the operator is added, so the scissor Hamiltonian related to those terms are already dealt with implicitly. The second term of eq. (7.10), however, will need to be dealt explicitly for the scissor Hamiltonian. We will once again separate the Hamiltonian as $H_{\beta\alpha} = H^{\mathrm{o}}_{\beta\alpha} + H^{\mathrm{scissor}}_{\beta\alpha}$ and consider only the derivative of $H^{\mathrm{scissor}}_{\beta\alpha}$, which is

$$\begin{aligned}
K_n^{\alpha\beta}\frac{\partial H^{\mathrm{scissor}}_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} &= K_n^{\alpha\beta}\frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})}S_{\beta\eta}K^{\eta\delta}_{\mathrm{shifted}}S_{\delta\alpha} \\
&= K_n^{\alpha\beta}\frac{\partial S_{\beta\eta}}{\partial \phi^*_\alpha(\mathbf{r})}K^{\eta\delta}_{\mathrm{shifted}}S_{\delta\alpha} + K_n^{\alpha\beta}S_{\beta\eta}\frac{\partial K^{\eta\delta}_{\mathrm{shifted}}S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \\
&= K^{\eta\delta}_{\mathrm{shifted}}S_{\delta\alpha}K_n^{\alpha\beta}\frac{\partial S_{\beta\eta}}{\partial \phi^*_\alpha(\mathbf{r})} + K_n^{\alpha\beta}S_{\beta\eta}\frac{\partial K^{\eta\delta}_{\mathrm{shifted}}S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})}.
\end{aligned} \tag{7.12}$$

The first term here can simply be added to the first term of eq. (7.10). The second term is more complicated, as the derivative of $K^{\eta\delta}_{\text{shifted}}$ contains the derivative of an inverse overlap matrix. We can get around this by writing out $K^{\eta\delta}_{\text{shifted}}$ in full (eq. (7.4)), so that the inverse overlap matrix is removed by the overlap matrix:

$$\begin{aligned} K_n^{\alpha\beta} S_{\beta\eta} \frac{\partial K^{\eta\delta}_{\text{shifted}} S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} &= K_n^{\alpha\beta} S_{\beta\eta} \frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})} \sum_L \left(\sigma_v^L K_L^{\eta\delta} + \sigma_c^L (S_L^{\eta\delta} - K_L^{\eta\delta}) \right) S_{\delta\alpha} \\ &= K_n^{\alpha\beta} S_{\beta\eta} \frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})} \sum_L \left(\sigma_v^L K_L^{\eta\delta} S_{\delta\alpha} + \sigma_c^L S_L^{\eta\delta} S_{\delta\alpha} - \sigma_c^L K_L^{\eta\delta} S_{\delta\alpha} \right) \\ &= K_n^{\alpha\beta} S_{\beta\eta} \frac{\partial}{\partial \phi^*_\alpha(\mathbf{r})} \sum_L \left((\sigma_v^L - \sigma_c^L) K_L^{\eta\delta} S_{\delta\alpha} + \sigma_c^L \delta^\eta{}_\alpha \right) \\ &= K_n^{\alpha\beta} S_{\beta\eta} \sum_L (\sigma_v^L - \sigma_c^L) \left(\frac{\partial K_L^{\eta\delta}}{\partial \phi^*_\alpha(\mathbf{r})} S_{\delta\alpha} + K_L^{\eta\delta} \frac{\partial S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \right). \end{aligned} \tag{7.13}$$

Using the purification transformation and property of the derivative from eq. (7.9), this derivative turns out to be zero:

$$\begin{aligned} \frac{\partial K^{\eta\delta}_{\text{shifted}} S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} &= \sum_L \Delta_L \Big((3LSL - 2LSLSL - 2LSLSL)_L^{\eta\delta} \frac{\partial S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \\ &\qquad + (3LSL - 2LSLSL)_L^{\eta\delta} \frac{\partial S_{\delta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \Big) \\ &= 0 \end{aligned} \tag{7.14}$$

Hence, the energy derivative is

$$\begin{aligned} \frac{\partial E}{\partial \phi^*_\alpha(\mathbf{r})} &= \left(Q^{\alpha\beta} + (K_{\text{shifted}} S K_n)^{\alpha\beta} \right) \frac{\partial S_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} + K_n^{\alpha\beta} \frac{\partial H^{\text{o}}_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} \\ &= \left(Q^{\alpha\beta} + (K_{\text{shifted}} S K_n)^{\alpha\beta} \right) \left(\phi_\beta + \sum_{ij} \tilde{p}^i(\mathbf{r}) O_{ij} \left\langle \tilde{p}^j \middle| \phi_\alpha \right\rangle \right) + K_n^{\alpha\beta} [H^{\text{o}} \phi_\beta](\mathbf{r}), \end{aligned} \tag{7.15}$$

with

$$\frac{\partial S_{\beta\alpha}}{\partial \phi^*_\alpha(\mathbf{r})} = \phi_\beta + \sum_{ij} \tilde{p}^i(\mathbf{r}) O_{ij} \left\langle \tilde{p}^j \middle| \phi_\alpha \right\rangle, \tag{7.16}$$

because of the PAW transformation described in section 3.5.

In summary, the species dependent scissor operator in eq. (7.3) has been added, and the $K_{\text{shifted}} S K_n$ term in eq. (7.15) has been implemented to the NGWF gradient calculation in ONETEP. The results section of this chapter will test several

different options for how the scissor operator can be used.

7.3 Strain Analysis of MoS_2 and TiS_2

As discussed earlier, fig. 7.1a shows several pairs of MS_2 materials where DFT and GW predicts different types of band alignment. For this project, we have chosen to study MoS_2 and TiS_2. This choice is motivated by a need to test the validity of our method on an exemplar system, and monolayer MoS_2 has already been well explored. From fig. 7.1b, we can see that for MoS_2 and TiS_2, DFT predicts a broken band gap, but GW predicts a staggered band gap with a band gap of 0.68 eV. Before we simulate a large supercell of MoS_2/TiS_2 VHS, we first demonstrate why one cannot simply strain both materials to the same lattice constant so that a supercell is not necessary.

The DFT calculations in this section were made using the Quantum Espresso package [44]. We employ the PAW formalism [71] with the scalar relativistic pseudopotentials with the PBE functional [33] from the JTH PAW atomic datasets [175]. SOC was not included because we just wanted to see how the bands move as a function of strain, so the spin-orbit splitting is not important in this test. For the S, Ti and Mo atoms, the outermost 6, 12 and 14 electrons were retained as valence. 80 Ry was used for the plane-wave cutoff energy.

The MoS_2 and TiS_2 structures were first optimised using geometry optimisation with cell relaxation until forces were smaller than $10^{-4}\,\mathrm{Ry\,Bohr^{-1}}$. This results in a lattice constant of 3.19 Å and 3.34 Å for MoS_2 and TiS_2, respectively. Geometry optimisations and the self-consistent calculations were performed with an 18×18 in-plane k-point grid. To avoid interaction between the out-of-plane periodic images (i.e., the unphysical interaction between the repeated monolayers from the periodic boundary condition), the vacuum spacing was 25 Å.

Four band structures were calculated for MoS_2 and TiS_2, with lattice parameters ranging from the optimal lattice constant to the average lattice constant 3.26 Å. The band structures are shown in fig. 7.2 as a gradient from red to yellow. As can be seen, the effect of strain on both material is clearly not negligible. The MoS_2 transitions from a direct band gap to an indirect band gap. In addition, the band gap shrinks as strain increases. The TiS_2 on the other hand maintains the band gap, but with both the VBE and the CBE shifted up, which in fact changes their band alignment to a staggered gap. Therefore, applying the strain to both material would change their properties drastically. Running a GW calculation on a small unit cell of TiS_2/MoS_2 VHS with lattice constant 3.26 Å would hence yield

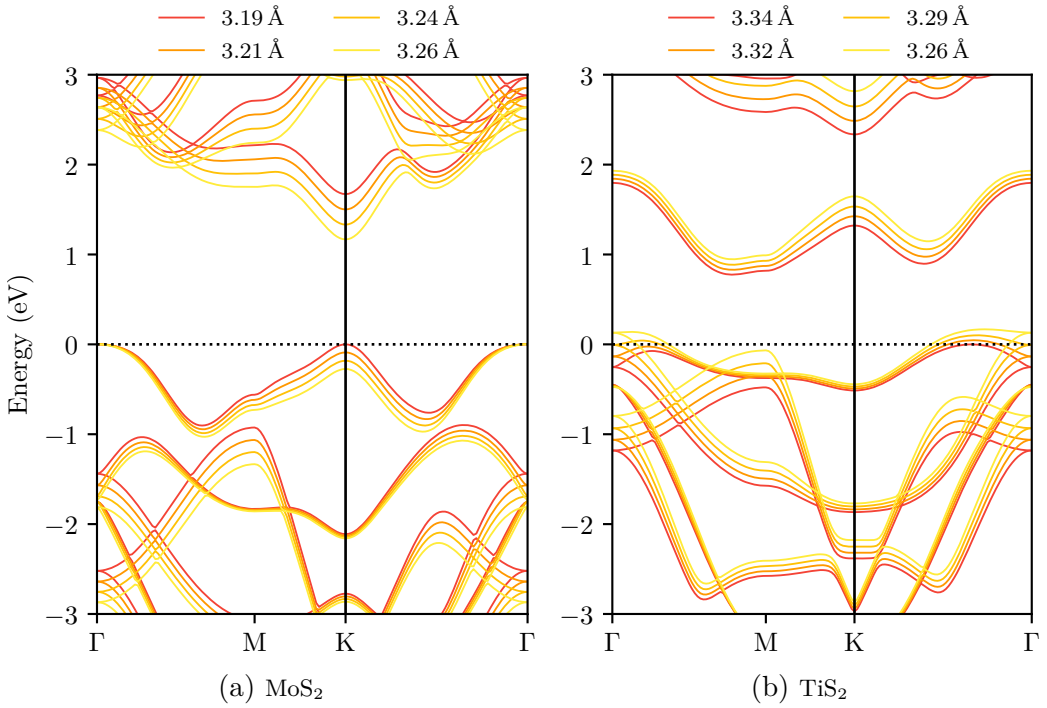


FIGURE 7.2: DFT band structures of (a) MoS_2 with increasing lattice constant from the optimal lattice constant 3.19 Å to 3.26 Å and (b) TiS_2 with decreasing lattice constant from the optimal lattice constant 3.34 Å to 3.26 Å. The four band structure calculations are displayed as a gradient from red to yellow.

results that are not useful in understanding electronic properties.

7.4 Construction of Supercell

It has been established that we need to use a large supercell in order to capture the correct properties of an unstrained MoS_2/TiS_2 VHS. This section will discuss the construction of our supercell, as it is not intuitive. If we consider a VHS consisting of a substrate that is the bottom layer and an overlayer on top of the substrate, one can calculate the number of unit cells required to minimise the strain of the two layers by solving

$$a_{\text{substrate}} n = a_{\text{overlayer}} m, \tag{7.17}$$

where $a_{\text{substrate}}$ and $a_{\text{overlayer}}$ are the lattice constant of the substrate and the overlayer, while n and m are to be determined integers which gives the number of unit cells required for the substrate and the overlayer. For the lattice constants of 3.19 Å and 3.34 Å for MoS_2 and TiS_2, n and m would be unreasonably high (e.g., $n = 319$ and $m = 334$) and is infeasible for a DFT calculation. Thus, for our purposes we

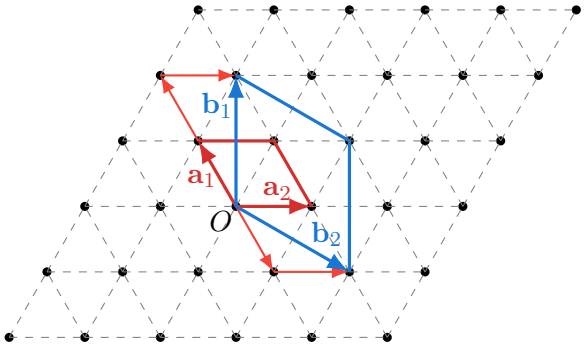


FIGURE 7.3: Examples of non-trivial hexagonal surface supercells. O denotes the origin of the grid. The red cell is a primitive cell with lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ shown with arrows. The blue cell is a $(\sqrt{3} \times \sqrt{3})$ cell rotated by 30° with lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. The thin red arrows illustrate the path needed to construct the blue supercell from the primitive cell.

have adopted a searching algorithm that searches different substrate supercells then find a matching overlayer supercell. We allow small amount of strain on the overlayer, and a small rotation is also included to help relax the strain and reduce the supercell size.

We search the substrate supercell by using a linear transformation on a primitive cell lattice vectors matrix. That is,

$$\mathbf{M}\mathbf{A}_{\text{old}} = \begin{pmatrix} n & m \\ k & l \end{pmatrix} \begin{pmatrix} \mathbf{a}_1^{\text{old}} \\ \mathbf{a}_2^{\text{old}} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1^{\text{new}} \\ \mathbf{a}_2^{\text{new}} \end{pmatrix} = \mathbf{A}_{\text{new}}, \tag{7.18}$$

where n, m, k and l are integers. As an example on how this works, fig. 7.3 illustrates a set of primitive cell lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ mapped to $\mathbf{b}_1$ and $\mathbf{b}_2$ after a transformation. In this example, we can see that to get from O to $\mathbf{b}_1$, we can go two steps $\mathbf{a}_1$ and one step $\mathbf{a}_2$. Meanwhile to get from O to $\mathbf{b}_2$, we can go one step $-\mathbf{a}_1$ and one step $\mathbf{a}_2$. Hence we can write the transformation

$$\mathbf{M}\mathbf{A} = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} 2\mathbf{a}_1 + \mathbf{a}_2 \\ -\mathbf{a}_1 + \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{B}. \tag{7.19}$$

Using this formulation, we can obtain the number of atoms in the new supercell by the determinant of $\mathbf{M}$ multiplied by the number of atoms in the primitive cell.

As one can imagine that many configurations of supercell can be searched through by using a transformation matrix. We simply generate many combinations of integers for n, m, k and l for the substrate, then we find the corresponding

transformation for the overlayer by solving the following equation

$$\mathbf{MA} = \begin{pmatrix} n & m \\ k & l \end{pmatrix} \begin{pmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} p & q \\ r & s \end{pmatrix} \begin{pmatrix} \lambda_1 & \sigma_1 \\ \sigma_2 & \lambda_2 \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{PSB}. \tag{7.20}$$

$\mathbf{B}$ is the rotated primitive lattice vectors matrix of the overlayer, in which $\mathbf{b}_1$ and $\mathbf{b}_2$ are rotated by angle θ via a rotation matrix such that

$$\mathbf{b}_1 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mathbf{b}_1^{\text{old}} \quad \text{and} \quad \mathbf{b}_2 = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mathbf{b}_2^{\text{old}}; \tag{7.21}$$

the matrix $\mathbf{P}$ is similar to $\mathbf{M}$, where it transforms a set of primitive lattice vectors to a create new supercell, thus the matrix elements p, q, r, s are integers, and $\mathbf{S}$ is a strain and shear transformation. We can obtain these two matrices by multiplying both side by the inverse of $\mathbf{B}$, giving

$$\mathbf{PS} = \mathbf{MAB}^{-1}. \tag{7.22}$$

As $\mathbf{P}$ is an integer matrix, we simply round the $\mathbf{MAB}^{-1}$ matrix to get $\mathbf{P}$. That is,

$$\mathbf{P} = \lfloor \mathbf{MAB}^{-1} \rceil. \tag{7.23}$$

$\mathbf{S}$ is then simply

$$\mathbf{S} = \mathbf{P}^{-1}\mathbf{MAB}^{-1}. \tag{7.24}$$

As one can see here that $\mathbf{S}$ is close to the identity matrix as it is just a small error introduced by the rounding procedure. After we obtain all the aforementioned matrices, we select as the best configuration the one with the lowest amount of strain and shear, while having a minimum number of atoms per layer of 200 and a maximum number of atoms per layer of 600. The minimum number of atoms is set so that a supercell size will be large enough to hold the NGWFs in ONETEP and allow for a large enough cell to sample higher number of k-points, as ONETEP samples only the Γ-point.

In summary, we search through a set of angles θ for the overlayer, which for our MoS_2/TiS_2 VHS, we searched through 2–6° at an interval of 0.1°. For each θ, we then generate all combinations of n, m, k and l, which we used the range of integers 0 to 12, so that the largest supercell we can make is a 12×12 supercell. The relaxed lattice constants from last section were used, where MoS_2 is 3.19 Å and TiS_2 is 3.34 Å. Afterward, we compute $\mathbf{P}$ using eq. (7.22) and $\mathbf{S}$ using eq. (7.24). We then rank each supercell configuration by the sum of the absolute value of the

elements in $(\mathbf{S} - \mathbf{I})$, where $\mathbf{I}$ is the identity matrix, and the lowest value of the sum being the highest rank. This procedure gives us $\theta = 4.3°$ and the following matrices:

$$\mathbf{M} = \begin{pmatrix} 12 & 1 \\ 11 & 12 \end{pmatrix}, \quad \mathbf{P} = \begin{pmatrix} 11 & 0 \\ 11 & 11 \end{pmatrix} \quad \text{and} \quad \mathbf{S} = \begin{pmatrix} 0.999\,97 & 0.000\,13 \\ -0.000\,13 & 0.999\,83 \end{pmatrix}.$$

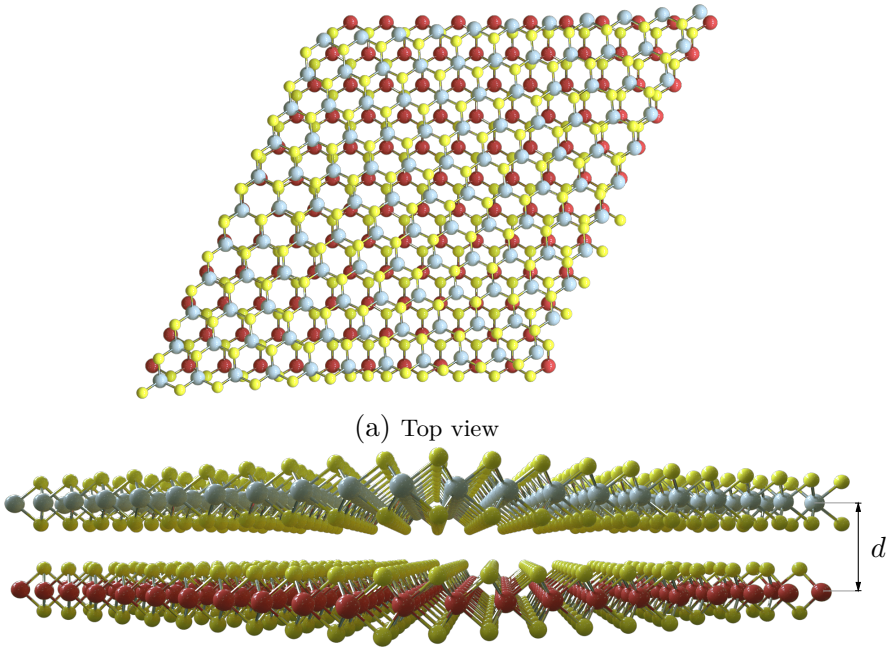
Our MoS_2/TiS_2 VHS therefore contain 762 atoms with the substrate (MoS_2) containing 399 atoms and the overlayer (TiS_2) containing 363 atoms. As we can see that the strain and shear is negligible and the rotation angle is very small, so this is a very good approximation to the unstrained VHS. The resulting supercell is shown in fig. 7.4. A 2×2 version of the supercell is shown in fig. 7.5, which clearly displays Moiré patterns with aligned and non-aligned regions.

7.5 Validation Tests With MoS_2

In this section, MoS_2 monolayer calculations using the scissor operator will be presented. The calculations were performed using the supercell discussed in the last section, but include only the MoS_2. The calculations therefore serve a threefold purpose: test whether the supercell created in the last section is indeed periodic and produces correct band structure, the function of the scissor operator, and finally the total energy can be used for the interlayer distance optimisation, which will be described in the next section.

The total energy calculations were performed using ONETEP [61]. The optB88-vdW XC functional was used for all our calculations, as it has shown to be the best option for a similar VHS [173]. It was also used for the monolayers to keep the XC functional consistent across all of our calculations. 1200 eV plane wave kinetic energy cutoff was chosen and we employed the PAW formalism [71, 73]. The NGWF radius was 13 Bohr for all the atoms. The same scalar relativistic pseudopotentials from the JTH PAW atomic datasets [175] were used as section 7.3. To avoid interaction between periodic images, the cell size in the non-periodic direction was 29.4 Å. A smearing of 0.1 eV was used for the DOS calculations.

We obtain the shifts needed to reproduce the GW band gap from [159], which is in a previously shown fig. 7.1b. For MoS_2, the valence band shift is -0.0733 eV and the conduction band shift is 0.8792 eV. The scissor operator is setup according to the shifts. The DOS and spectral function of the MoS_2 monolayer, with and without the scissor operator applied, are shown in fig. 7.6 and fig. 7.7. The band gap without the scissor operator is 1.66 eV, which is somewhat higher than the reported DFT values of 1.58 eV [159, 176]. However, this can be seen to be due to



(a) Top view

(b) Side view

FIGURE 7.4: A MoS_2/TiS_2 VHS supercell, where the bottom layer is MoS_2 and the top layer is TiS_2. The red atoms are Mo, the grey atoms are Ti and the yellow atoms are S. The interlayer distance d is defined by the distance between the metal atoms of each layer.

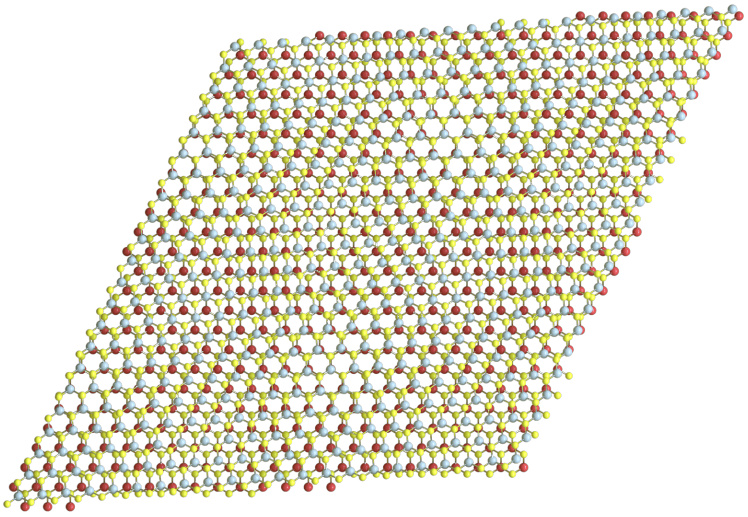


FIGURE 7.5: 2×2 MoS_2/TiS_2 VHS showing Moiré patterns containing aligned and non-aligned regions

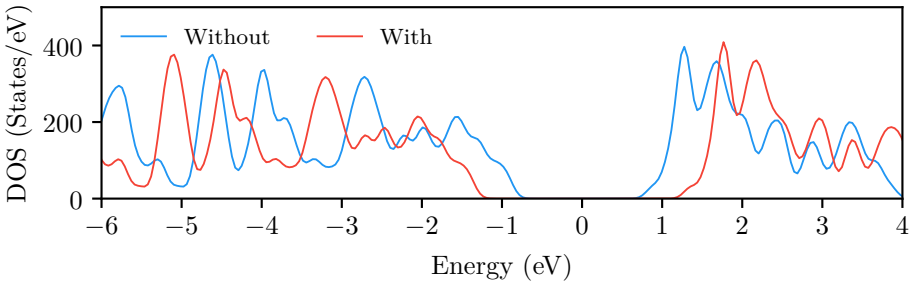


FIGURE 7.6: MoS_2 monolayer DOS comparing with and without scissor operator. The Fermi level is pinned at the middle the band gap. The with scissor LDOS shows a simple opening of the band gap with all the features of the without scissor LDOS left intact.

the lack of SOC. If we include the DFT spin-orbit splitting of 0.15 eV at K from section 6.9, hence decreasing the band gap by 0.075 eV, we obtain a band gap of ∼1.58 eV in accordance with literature results.

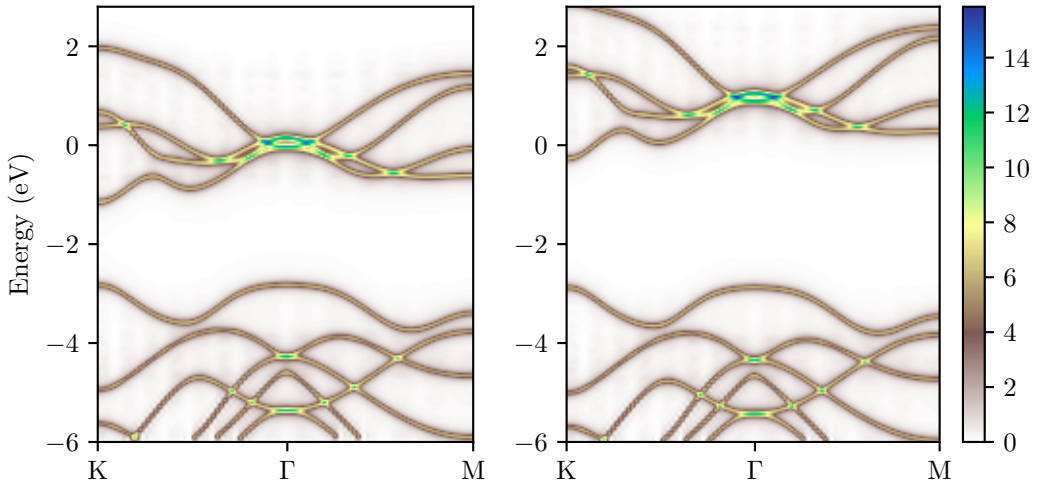
Unsurprisingly, both the DOS and spectral function with the scissor operator display a simple energy shift from the without-scissor operator results. The band gap increased to 2.61 with the scissor operator applied, an increase of 0.95 eV, which is expected from the shifts applied. Hence, we have demonstrated that the scissor operator has been implemented correctly when applied globally. In the next two sections, we will finally investigate the validity of using the scissor operator on the MoS_2/TiS_2 VHS.

7.6 Interlayer Distance and the Effects of the Scissor Operator

We have discussed the construction of our supercell previously but have left out a very important part of the design of our VHS, the interlayer distance. The interlayer distance d is the separation distance between the two layers, as illustrated in fig. 7.4. The optimal interlayer distance can be found by minimising the binding energy as a function of d. We compute the binding energy using

$$E_{\mathrm{b}} = E_{\mathrm{MoS_2/TiS_2}} - (E_{\mathrm{MoS_2}} + E_{\mathrm{TiS_2}}), \tag{7.25}$$

where $E_{\mathrm{MoS_2/TiS_2}}$ is the total energy of the MoS_2/TiS_2 VHS, $E_{\mathrm{MoS_2}}$ and $E_{\mathrm{TiS_2}}$ are the total energy of the isolated MoS_2 monolayer and TiS_2 monolayer that makes up the VHS.



(a) Without scissor (b) With scissor

FIGURE 7.7: MoS_2 monolayer spectral function (a) without scissor operator and (b) with scissor operator. The with scissor spectral function shows a simple opening of the band gap with all the features of the without scissor bands left intact.

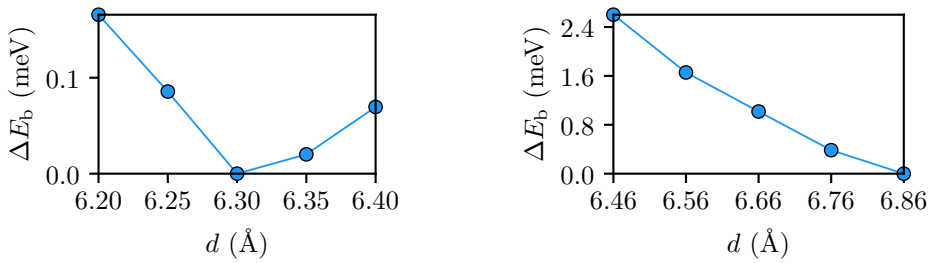
The ONETEP calculations set up is same as the last section with the addition of the EDFT functionality within ONETEP [177], because of the predicted broken band gap, thus the metallic nature of the MoS_2/TiS_2 VHS. See table 7.1 for a summary of all the parameters used. Geometry optimisation was not performed due to the large computational cost associated with such a large system, so that both layers are just kept flat. Because of this, it is important to note that our system is only an approximation, as twisted bilayers tend to buckle, where the aligned regions tend to buckle away from each other and the non-aligned regions tend to buckle towards each other.

Since no known studies have been found for MoS_2/TiS_2 VHS, we had no reference points for the interlayer distance. However, we can deduce that the interlayer distance is between 6.2 Å and 6.6 Å, because of the reported interlayer distances of 6.21 Å and 6.65 Å for MoS_2/WS_2 [178] and $MoS_2/MoSe_2$ [173], respectively. We expect the interlayer distance to be higher than that of MoS_2/WS_2, because TiS_2 has a higher lattice constant than WS_2. Meanwhile, we expect the interlayer distance to be lower than that of $MoS_2/MoSe_2$, because of the inclusion of Se tends to increase the thickness of the monolayer. We therefore calculated the binding energy from 6.20 Å to 6.40 Å interlayer distances at an interval of 0.05 Å.

The variation of binding energy with interlayer distance d is shown in fig. 7.8a. The optimal value was found to be 6.30 Å, which is in the aforementioned expected range. Note that one could perform a finer-grain search at a smaller d interval to

Parameter	Value
Plane-wave cutoff energy	1200 eV
NGWF radius	13 Bohr
EDFT	True
PAW	True
XC functional	optB88-vdW
DOS smearing	0.1 eV
MoS_2 Valence band shift	−0.0733 eV
MoS_2 Conduction band shift	0.8792 eV
TiS_2 Valence band shift	−0.7217 eV
TiS_2 Conduction band shift	0.7516 eV

TABLE 7.1: A table showing several of the most important parameters used in ONETEP, which would be necessary to reproduce the calculations performed.



(a) Without scissor (b) With scissor

FIGURE 7.8: Interlayer distance d optimization, with the lowest binding energy taken as energy reference, where (a) is without the scissor operator and (b) is with the scissor operator applied.

obtain a more accurate result, but for our purposes of validating the scissor operator functionality, this accuracy is enough. The scissor operator was not applied to obtain the above interlayer distance, but it was particularly interesting to see what would happen if we do apply the scissor operator.

We obtain the shifts needed to reproduce the GW band alignment from [159], results from when were previously shown as fig. 7.1b. For MoS_2, the valence band shift is $-0.0733\,\text{eV}$ and conduction band shift is $0.8792\,\text{eV}$, as previously. For TiS_2, the valence band shift is $-0.7217\,\text{eV}$ and conduction band shift is $0.7516\,\text{eV}$. The scissor operator is setup according to the shifts, and the resulting binding energy as a function of interlayer distance is shown in fig. 7.8b. It shows that the binding energy continue to decrease even for much higher d than the optimal value predicted by calculations without the scissor operator. This is likely to be attributed to unphysical effects introduced by the scissor operator.

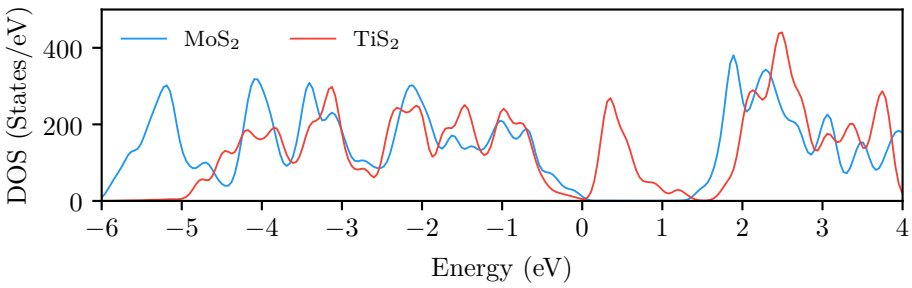
One can interpret the scissor operator as applying an external electric field to the system, similar to that of applying an electrostatic gate voltage, which causes the band gap renormalisation. However in this case, there would be two separate non-interacting electric fields acting on the individual layers with no self-consistent polarisation response, which needless to say is unphysical and causes the two layers to separate. Therefore it is recommended to not use the scissor operator for interlayer distance optimisation.

7.7 Effects of the Scissor Operator on LDOS and Spectral Function

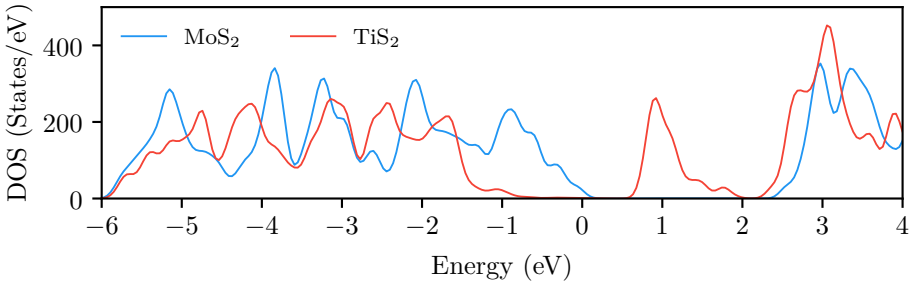
7.7.1 LDOS Analysis

Finally, with the interlayer distance determined, the construction of the MoS_2/TiS_2 supercell is complete, and we can begin some properties calculations related to the scissor operator. We first investigated the LDOS of the individual MoS_2 layer and TiS_2 layer comparing results without and with the scissor operator. These are shown in fig. 7.9a and fig. 7.9b.

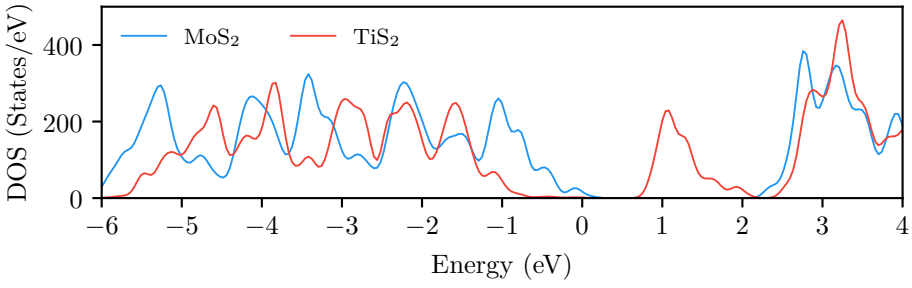
It can be seen that the band gap opened for both MoS_2 and TiS_2. The MoS_2 local band gap without and with scissor is $1.46\,\text{eV}$ and $2.42\,\text{eV}$, respectively. Therefore, the scissor operator increased the gap by $0.96\,\text{eV}$, which differs by $0.01\,\text{eV}$ from the expected $0.95\,\text{eV}$. The expected band gap increased is simply from the magnitude of the shifts applied. Interestingly, the local band gap of TiS_2 without scissor is a mere $0.30\,\text{eV}$, much lower than the reported $0.73\,\text{eV}$ [159] and our own



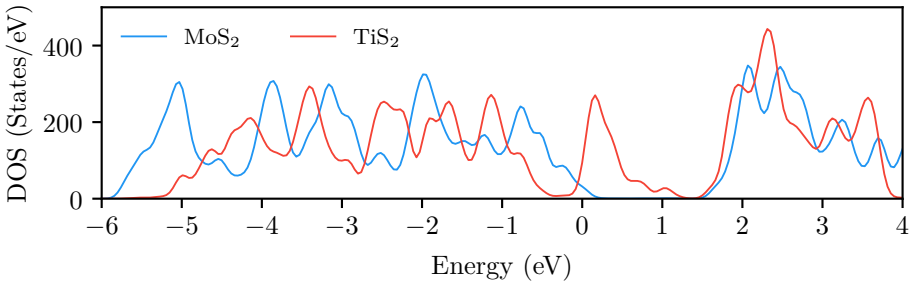
(a) Without scissor



(b) With scissor



(c) Without→with scissor



(d) With→without scissor

FIGURE 7.9: MoS_2/TiS_2 VHS LDOS comparing different ways to use the scissor operator.

calculation without SOC 0.80 eV. Moreover, the band gap of the VHS opened to 0.77 eV, which is above the predicted 0.68 eV from fig. 7.1b. Upon closer inspection, we can see that even without the scissor operator, the VHS has a band gap of 0.30 eV which seems to comes from just the TiS_2. The scissor operator has therefore only opened the band gap by 0.47 eV.

We first thought that these 'strange' behaviours were caused by the use of the scissor operator for the full SCF calculation, as we saw that the scissor operator can produce incorrect results for the interlayer distance optimisation. Hence, we examined whether one can run the scissor operator after a self-consistent calculation without the scissor operator. This process of first running without scissor then with scissor will be denoted as 'without→with scissor' from now on. Furthermore, to see how running full SCF with scissor can affect the results, we have also tested running first with scissor, then without (this will henceforth be denoted as 'with→without scissor') to see whether we will retrieve the original without scissor DOS.

The LDOS for the without→with scissor is illustrated in fig. 7.9c. It shows that, once again, the band gap for both MoS_2 and TiS_2 have increased. However, for MoS_2, it only increased to 2.14 eV, which is far from the expected value. Unfortunately, the band gap of TiS_2 is hard to measure, as we shall see why later from their spectral functions. The band gap of the VHS is the same as with scissor. This is caused by the TiS_2 conduction states being shifted to a higher energy relative to the with scissor case. By also looking at the with→without scissor LDOS shown in fig. 7.9d, we can see running full SCF calculations with scissor indeed greatly affects the electronic structures as one cannot reproduce a without scissor results afterward, while re-using the density and NGWFs.

Within ONETEP, a full SCF calculation optimises two nested loops the density kernel and the NGWFs. One can run a full SCF calculation without scissor to optimise the NGWFs and density kernel, then optimise only the density kernel with the scissor operator. On the contrary, one cannot optimise solely the NGWFs with the scissor operator, because the optimisation of NGWFs requires convergence of the density kernel. We have investigated whether it is possible to only optimise the density kernel after a full SCF without scissor to produce the same result as the full SCF with scissor. The LDOS of such case shown in fig. 7.10. The LDOS is very similar to with scissor (fig. 7.9b). The band gap, however, is 0.75 eV, smaller than with scissor 0.77 eV. All these initial findings indicate that running a full SCF with scissor is perhaps necessary if one wishes to use the scissor operator, but just a density kernel optimisation with scissor after a full SCF without scissor seems possible.

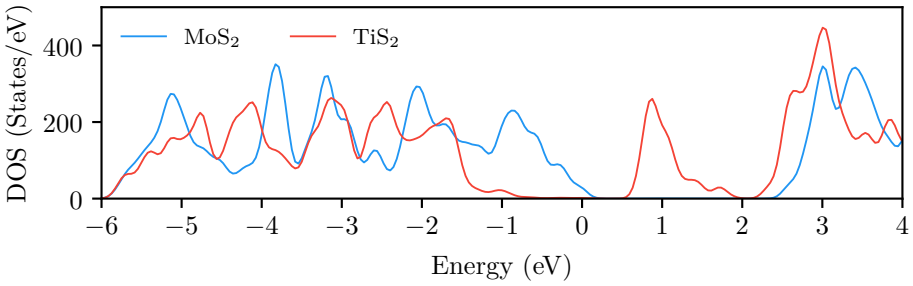


FIGURE 7.10: MoS_2/TiS_2 VHS LDOS for a full SCF without scissor operator followed by a density kernel optimisation with scissor operator.

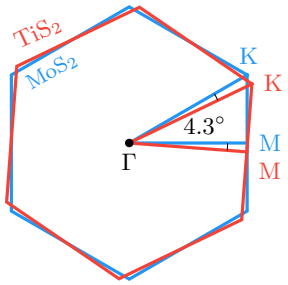


FIGURE 7.11: Schematic of overlapping monolayer Brillouin zone, with the TiS_2 layer (in red) rotated by 4.3° relative to the MoS_2 layer (in blue).

7.7.2 Spectral Function Analysis

To explain all the aforementioned behaviours, we turn to the spectral functions. We calculate the spectral function by unfolding the single-particle supercell spectral function to the monolayer of interest [173]. With this method, we can analyse the spectral function of the MoS_2 layer and the TiS_2 layer separately, similar to the LDOS. Since the two layers have two different BZ, with TiS_2 rotated by 4.3° relative to MoS_2, the high symmetry points are also different, see fig. 7.11. We therefore colour code the high symmetry points with blue for when the spectral function is sampled in the MoS_2 BZ, and red for when the spectral function is sampled in the TiS_2 BZ. In order the provide a reference to how the bands of each layer changed, the spectral functions for the MoS_2 and TiS_2 monolayer are shown in fig. 7.12. Only the path K-Γ-M is computed, because we expect the VBE to be between Γ to K and the CBE to be between Γ to M (from our strain analysis, fig. 7.2).

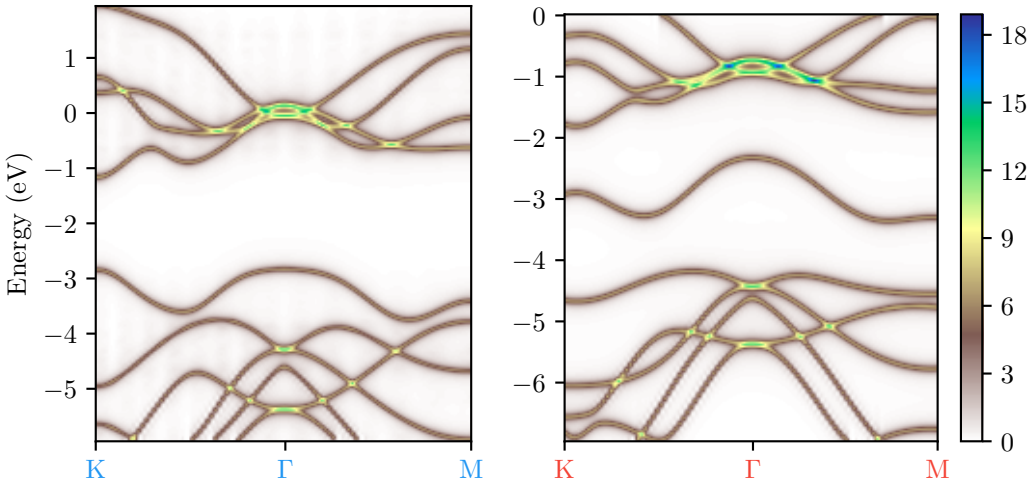
The spectral functions for without scissor, with scissor and without→with scissor are shown in fig. 7.13, fig. 7.14 and fig. 7.15. The first problem we faced was the small band gap of TiS_2 when calculated without scissor. Figure 7.13 shows the hybridisation of the HOMO band (i.e., the MoS_2 HOMO band appearing in the TiS_2

spectral function), thus reducing the band gap to be between the MoS_2 VBE to the TiS_2 CBE. In fact, most of the high spectral weight bands near Γ in TiS_2 are from the MoS_2 bands. These hybridisations seem unlikely to be physical, and the scissor operator appears to be able to correct these. Figure 7.14 shows the reduction of the spectral weight of the MoS_2 bands in TiS_2. However, experimental measurements are needed to confirm whether the hybridisations are real physical phenomena.

The second problem observed was that the existence of the band gap even without scissor. As can be seen that the band gap of MoS_2 is indirect. However, this is expected, as the bilayer MoS_2 also possesses an indirect band gap [176], where the bands at K get shifted down, thus causing the VBE to be at Γ. This explains why the band gap exists, because the predicted overlap between the TiS_2 LUMO band and the MoS_2 HOMO band is only 0.14 eV. The lack of SOC would have already reduced the overlap, so when the VBE transitions from K to Γ, it would create a small gap in the VHS. In addition, it explains the decrease in the MoS_2 local band gap in our VHS.

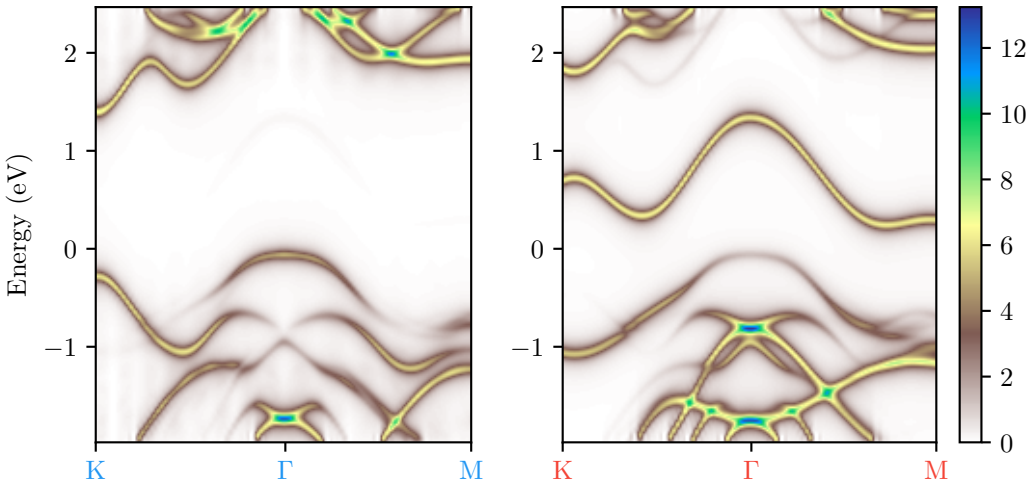
The third problem noted was that the scissor operator was only able to open the band gap of the VHS by 0.47 eV. We can see by comparing fig. 7.13 and fig. 7.14, that the LUMO band of TiS_2 appears not to have been shifted up by the intended amount of 0.75 eV. Unfortunately due the hybridisation, it is too difficult to identify the without scissor band gap associated with TiS_2, but examining the with scissor band gap of 2.34 eV, the scissor operator has certainly corrected the gap to some extent. Hence, the problem lies in the incorrect band alignment, as both MoS_2 and TiS_2 seems to show the correct band gap opening by the scissor operator. There is one shortcoming in our implementation that perhaps explains this. The operator only shifts the selected layers, the 'cross-terms' are left untouched. For example, the terms in the density kernel that relates the Mo and Ti NGWFs are not shifted.

Finally, we discussed whether one can perform a without→with scissor calculation. It can be seen from fig. 7.15, the shape of the MoS_2 HOMO band has changed drastically. This would suggest that the scissor operator decreased the effective mass at Γ, which, needless to say, is erroneous. Not to mention the wrong band gap opening that was established in the LDOS discussion. Hence, it is not recommended to only apply the scissor operator non-self-consistently after a full SCF calculation without scissor.



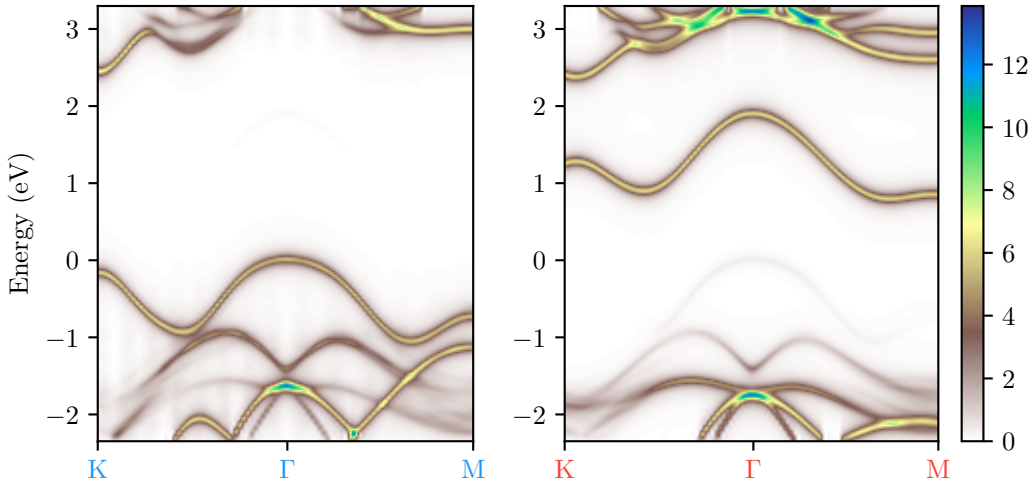
(a) MoS_2 (b) TiS_2

FIGURE 7.12: Spectral functions of (a) MoS_2 monolayer and (b) TiS_2 monolayer. It shows hybridisation of the HOMO band, where the MoS_2 HOMO band appearing in the TiS_2 spectral function.



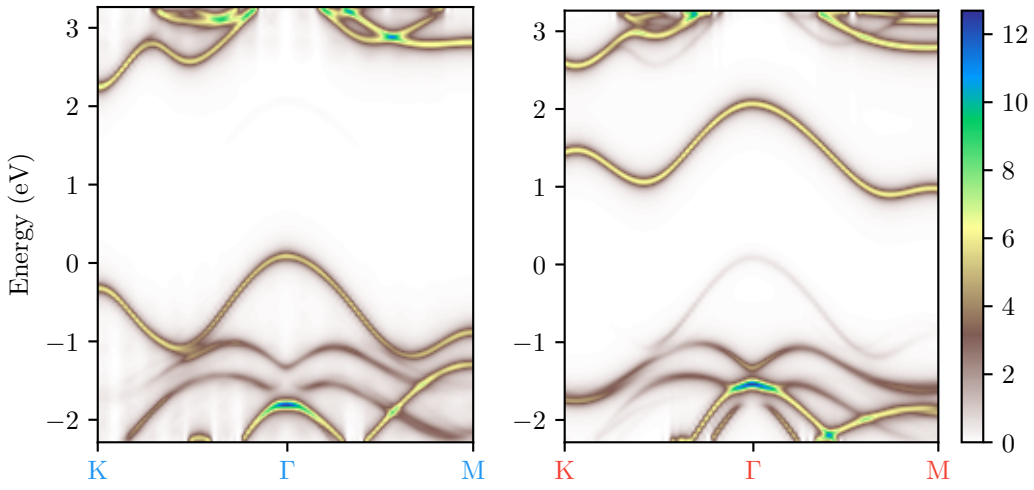
(a) MoS_2 (b) TiS_2

FIGURE 7.13: Spectral function without scissor. It shows hybridisation of the HOMO band, and high spectral weight bands near Γ from the MoS_2 bands in the TiS_2.



(a) MoS_2 (b) TiS_2

FIGURE 7.14: Spectral function with scissor. It shows a reduction of the spectral weight intrusion from the MoS_2 bands in the TiS_2 spectral function compared to the without scissor spectral function in fig. 7.13.



(a) MoS_2 (b) TiS_2

FIGURE 7.15: Spectral function without→with scissor showing a large change in the shape of the HOMO band.

7.8 Conclusions

In conclusions, we have implemented a novel species-dependent scissor operator into the ONETEP code. The scissor operator method works by adding an extra potential term to the Hamiltonian that shifts the eigenvalues depending on the species. Because of the addition of the new Hamiltonian term, the NGWF gradient calculation was updated to reflect this change. When this species-dependent scissor operator is applied with a single shift to open up the band gap for the whole system, it becomes the traditional scissor operator, thus it can be used across any material with a band gap. However, for materials that require different shifts for different species, it is only recommended to be used between non-covalently-bonded layered materials, otherwise it would produce unphysical results. For example, one can use this method for VHSs but should not be used for LHSs or alloys.

Initial validation tests on MoS_2 monolayer proved that the functionality has been implemented correctly and can be used to increase the band gap of a material. The MoS_2/TiS_2 VHS was then investigated to demonstrate how the scissor operator can be used to correct the band alignment. MoS_2 and TiS_2 was chosen because of the predicted broken gap by DFT but GW predicts a type II band alignment. We have found that the scissor operator should not be used to perform interlayer distance optimisation for VHSs, as is causes the total energy to continue to decrease at high interlayer distance.

The MoS_2/TiS_2 VHS was found within plain DFT to have a small band gap of 0.3 eV, without the use of the scissor operator. This was attributed to the MoS_2 band gap becoming indirect when there are multiple vdW layers, and our lack of SOC in the calculation. It has been shown that one should not use the scissor operator after a full SCF calculation without it, otherwise it would produce incorrect band structure. The scissor operator has shown, not only to improve the band gap, but also correct the hybridisation of bands, though experimental work is needed to confirm this. Continued collaboration between this research group and that of Dr Neil Wilson is expected to lead to future ARPES experiments could be used to rigorously test this method. However, we have faced the problem that the band gap after the application of the scissor operator is not almost as large as expected, thus more tests are necessary before any real world use. Nevertheless, the species dependent scissor operator has shown promising results and have the potential for simulating large-scale VHSs with GW-like accuracy.

Chapter 8

Conclusion

In this thesis, we have focused on linear-scaling DFT for calculations on lateral and vertical heterostructures of TMDCs. We have developed approximations using DFT results to compute the critical thickness of LHSs and a framework for creating LHSs with periodic boundary condition for DFT calculations. A new algorithm was also developed and implemented into the linear-scaling DFT package ONETEP to better predict bandstructures for VHSs. Another area that was looked at was comparing GW and DFT bandstructures to experimental ARPES results. For the first time, direct measurement of the conduction band for several TMDCs monolayer, bilayer and trilayer using ARPES were obtained, and we were able to use GW approximation to compute quantitatively comparable bandstructures.

In chapter 5, we firstly explored the $WS_2|MoS_2$ LHS, we began by analysing the DOS and slabwise LDOS. The band gap was found to be 1.8 eV, while the LDOS shows the band alignment is a type II heterojunction, with the transition from bulk-like DOS of WS_2 to bulk-like DOS of MoS_2 takes place over just 1–2 slab widths. This small transition in the band alignment lead us to believe that the strong PL response at the interface shown from experiments could be due to the separation of defects. From our vacancy formation energy calculations as a function of position across the LHS, it was found that sulphur vacancies have a slight tendency to form at the W-Mo interface. Hence, we hypothesise that the enhanced PL observed experimentally could be due to the trapping of vacancies near one of the interfaces. Before we end the study of $WS_2|MoS_2$, we looked at the slight asymmetry in the vacancy formation energy and it was found to occur only in the zigzag interface due to the difference between the W-Mo interface and Mo-W interface.

We then explored the critical thickness of $MoS_2|MoSe_2$ LHS and have developed a simple approximation to compute the critical thickness of a thin film on a

thick substrate based on DFT calculations. In our model, we used a grain boundary nanoribbon that includes a dislocation and a rotated nanoribbon to extract the energy associated with adding a dislocation. By computing the energy per formula unit for a perfect film and a one-directional strained film, we can construct hypothetical systems that have different film thickness with a dislocation embedded and hence, calculate the critical thickness. We obtained the critical thickness of a thin MoS_2 film on a thick $MoSe_2$ substrate as 20.97 Å and 6.61 Å for a thin $MoSe_2$ film on a thick MoS_2 substrate. Despite the crude approximations, the critical thicknesses is in the expected range. As it would be interesting to use a full periodic LHS to compute the critical thickness, we ended the chapter with how one may design a full periodic LHS to be simulated using DFT.

In chapter 6, we compared bandstructures from DFT and GW with experimental ARPES measurements, and provided several suggestions on converging a GW calculation. We looked at WSe_2 and WS_2 1L, 2L and 3L and have shown that the GW band structures are in excellent agreement with ARPES, while DFT is only qualitatively comparable. The first problem we faced with the GW calculations was the bumpy numerical artefacts in the band structure and it was found that it is necessary to include the 3s and 3p states in Se. We therefore recommend one to check that certain semi-core states are included if faced with a similar problem. We have also demonstrated that the HF gap of 2D materials diverges for large k-point sampling size if the divergence of the Coulomb term is not properly dealt with, and recommend the use of the random integration method to deal with this divergence problem. We then looked at the problem that GW calculations tend to converge very slowly with the number of bands, and subsequently tested the effective energy technique (EET) and have shown better convergence even without increasing the number of bands, so for users with limited computing power, the EET can be used.

Looking into the future of large-scale DFT calculations on vertical heterostructures and the possible problem of incorrect band alignments predicted by DFT, we have presented, in chapter 7, a novel species-dependent scissor operator to solve such problem. The idea is to use band alignments from other more accurate methodologies such as GW approximation and apply those alignments to DFT. This new functionality has been implemented to ONETEP. The scissor operator method works by adding an extra potential term to the Hamiltonian that shifts the eigenvalues depending on the species, which in turns allow the band gap each layer in a VHS to be corrected individually. This method is only recommended to be used between non-covalently-bonded layered materials, otherwise it would produce unphysical results. In addition, a shortcoming of our scissor operator is that should

not be used to perform interlayer distance optimisation for VHSs, as is causes the total energy to continue to decrease at high interlayer distance.

We have applied the new scissor operator to the MoS_2/TiS_2 VHS, because of the predicted broken gap by DFT but GW predicts a type II band alignment, so this is a good demonstration of the problem it can solve. We have demonstrated that the scissor operator can not only to improve the band gap, but also correct the hybridisation of bands, though experimental work is needed to confirm this. One caveat is that the scissor operator cannot be applied after a full SCF calculation without it, otherwise it would produce incorrect band structure. More tests, including experimental work, are certainly needed to verify the accuracy of our method but it has shown some initial promising results and we believe that it can be very useful for multilayered VHSs with more complicated band alignments.

Bibliography

[1] P. V. Nguyen, N. C. Teutsch, N. P. Wilson, J. Kahn, X. Xia, A. J. Graham, V. Kandyba, A. Giampietri, A. Barinov, G. C. Constantinescu, N. Yeung, N. D. M. Hine, X. Xu, D. H. Cobden, and N. R. Wilson, "Visualizing electrostatic gating effects in two-dimensional heterostructures", Nature **572**, 220–223 (2019) (cit. on pp. ix, 57, 58, 62, 80).

[2] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim, "Two-dimensional atomic crystals", Proceedings of the National Academy of Sciences of the United States of America **102**, 10451–3 (2005) (cit. on p. 1).

[3] A. K. Geim and K. S. Novoselov, "The rise of graphene", Nature Materials **6**, 183–191 (2007) (cit. on p. 1).

[4] B. Dubertret, T. Heine, and M. Terrones, "The Rise of Two-Dimensional Materials", Accounts of Chemical Research **48**, 1–2 (2015) (cit. on pp. 1, 57).

[5] A. V. Kolobov and J. Tominaga, *Two-Dimensional Transition-Metal Dichalcogenides*, Vol. 239, Springer Series in Materials Science (Springer International Publishing, Cham, 2016) (cit. on p. 1).

[6] Y. Yoon, K. Ganapathi, and S. Salahuddin, "How Good Can Monolayer MoS2 Transistors Be?", Nano Letters **11**, 3768–3773 (2011) (cit. on p. 1).

[7] B. Radisavljevic, M. B. Whitwick, and A. Kis, "Integrated Circuits and Logic Operations Based on Single-Layer MoS2", ACS Nano **5**, 9934–9938 (2011) (cit. on p. 1).

[8] Y. Zhang, J. Ye, Y. Matsuhashi, and Y. Iwasa, "Ambipolar MoS2 Thin Flake Transistors", Nano Letters **12**, 1136–1140 (2012) (cit. on p. 1).

[9] K. F. Mak, K. He, J. Shan, and T. F. Heinz, "Control of valley polarization in monolayer MoS2 by optical helicity", Nature Nanotechnology **7**, 494–498 (2012) (cit. on p. 1).

[10] H. Zeng, J. Dai, W. Yao, D. Xiao, and X. Cui, "Valley polarization in MoS2 monolayers by optical pumping", Nature Nanotechnology **7**, 490–493 (2012) (cit. on p. 1).

[11] T. Cao, G. Wang, W. Han, H. Ye, C. Zhu, J. Shi, Q. Niu, P. Tan, E. Wang, B. Liu, and J. Feng, "Valley-selective circular dichroism of monolayer molybdenum disulphide", Nature Communications **3**, 887 (2012) (cit. on p. 1).

[12] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, "Emerging Photoluminescence in Monolayer MoS2", Nano Letters **10**, 1271–1275 (2010) (cit. on p. 1).

[13] G. Eda, H. Yamaguchi, D. Voiry, T. Fujita, M. Chen, and M. Chhowalla, "Photoluminescence from Chemically Exfoliated MoS 2", Nano Letters **11**, 5111–5116 (2011) (cit. on p. 1).

[14] Z. Yin, H. Li, H. Li, L. Jiang, Y. Shi, Y. Sun, G. Lu, Q. Zhang, X. Chen, and H. Zhang, "Single-Layer MoS2 Phototransistors", ACS Nano **6**, 74–80 (2012) (cit. on p. 1).

[15] B. V. Lotsch, "Vertical 2D Heterostructures", Annual Review of Materials Research **45**, 85–109 (2015) (cit. on pp. 1, 57).

[16] J. Wang, Z. Li, H. Chen, G. Deng, and X. Niu, "Recent Advances in 2D Lateral Heterostructures", Nano-Micro Letters **11**, 48 (2019) (cit. on p. 1).

[17] P. A. M. Dirac, "A new notation for quantum mechanics", Mathematical Proceedings of the Cambridge Philosophical Society **35**, 416 (1939) (cit. on p. 3).

[18] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln", Annalen der Physik **389**, 457–484 (1927) (cit. on p. 4).

[19] D. R. Hartree, "The Wave Mechanics of an Atom with a Non-Coulomb Central Field. Part I. Theory and Methods", Mathematical Proceedings of the Cambridge Philosophical Society **24**, 89 (1928) (cit. on p. 4).

[20] V. Fock, "Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems", Zeitschrift für Physik **61**, 126–148 (1930) (cit. on pp. 5, 69).

[21] P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas", Physical Review **136**, B864–B871 (1964) (cit. on p. 5).

[22] W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects", Physical Review **140**, A1133–A1138 (1965) (cit. on pp. 7, 10).

[23] F. Giustino, *Materials Modelling using Density Functional Theory* (Oxford University Press, 2014), pp. 41–43 (cit. on p. 8).

[24] R. G. Parr and W. Yang, *Density-functional theory of atoms and molecules* (Oxford University Press, 1989), p. 333 (cit. on pp. 10, 11).

[25] M. Gell-Mann and K. A. Brueckner, "Correlation Energy of an Electron Gas at High Density", Physical Review **106**, 364–368 (1957) (cit. on p. 10).

[26] O. Gunnarsson and B. I. Lundqvist, "Exchange and correlation in atoms, molecules, and solids by the spin-density-functional formalism", Physical Review B **13**, 4274–4298 (1976) (cit. on p. 11).

[27] E. Engel and R. M. Dreizler, *Density Functional Theory*, Theoretical and Mathematical Physics (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011) (cit. on p. 11).

[28] R. O. Jones and O. Gunnarsson, "The density functional formalism, its applications and prospects", Reviews of Modern Physics **61**, 689–746 (1989) (cit. on p. 11).

[29] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation", Physical Review B **46**, 6671–6687 (1992) (cit. on p. 11).

[30] J. P. Perdew, "Density-functional approximation for the correlation energy of the inhomogeneous electron gas", Physical Review B **33**, 8822–8824 (1986) (cit. on p. 11).

[31] J. P. Perdew and W. Yue, "Accurate and simple density functional for the electronic exchange energy: Generalized gradient approximation", Physical Review B **33**, 8800–8802 (1986) (cit. on p. 11).

[32] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior", Physical Review A **38**, 3098–3100 (1988) (cit. on p. 11).

[33] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple", Physical Review Letters **77**, 3865–3868 (1996) (cit. on pp. 11, 37, 65, 77, 90).

[34] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist, "Van der Waals Density Functional for General Geometries", Physical Review Letters **92**, 246401 (2004) (cit. on p. 11).

[35] D. C. Langreth and J. P. Perdew, “Exchange-correlation energy of a metallic surface: Wave-vector analysis”, Physical Review B **15**, 2884–2901 (1977) (cit. on p. 11).

[36] W. Kohn, Y. Meir, and D. E. Makarov, “van der Waals Energies in Density Functional Theory”, Physical Review Letters **80**, 4153–4156 (1998) (cit. on p. 11).

[37] G. Román-Pérez and J. M. Soler, “Efficient Implementation of a van der Waals Density Functional: Application to Double-Wall Carbon Nanotubes”, Physical Review Letters **103**, 096102 (2009) (cit. on pp. 11, 12).

[38] J. Klimeš, D. R. Bowler, and A. Michaelides, “Chemical accuracy for the van der Waals density functional”, Journal of Physics: Condensed Matter **22**, 022201 (2010) (cit. on pp. 12, 65, 77).

[39] J. P. Perdew, R. G. Parr, M. Levy, and J. L. Balduz, “Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy”, Physical Review Letters **49**, 1691–1694 (1982) (cit. on p. 13).

[40] L. J. Sham and M. Schlüter, “Density-Functional Theory of the Energy Gap”, Physical Review Letters **51**, 1888–1891 (1983) (cit. on p. 13).

[41] J. F. Janak, “Proof that dE/dn_i = epsilon in density-functional theory”, Physical Review B **18**, 7165–7168 (1978) (cit. on p. 13).

[42] C. Kittel, *Introduction to solid state physics* (Wiley, 2005), p. 680 (cit. on pp. 14, 47).

[43] K. N. Kirschner, J. B. Sorensen, and J. P. Bowen, “Calculating Interaction Energies Using First Principle Theories: Consideration of Basis Set Superposition Error and Fragment Relaxation”, Journal of Chemical Education **84**, 1225 (2007) (cit. on p. 15).

[44] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, “QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials”, Journal of Physics: Condensed Matter **21**, 395502 (2009) (cit. on pp. 15, 65, 90).

[45] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism", Physical Review B **41**, 7892–7895 (1990) (cit. on p. 16).

[46] R. M. Martin, *Electronic Structure* (Cambridge University Press, Cambridge, 2004) (cit. on p. 16).

[47] W. E. Pickett, "Pseudopotential methods in condensed matter applications", Computer Physics Reports **9**, 115–197 (1989) (cit. on p. 16).

[48] H. Hellmann, *Einführung in die quantenchemie* (F. Deuticke, Leipzig und Wien, 1937) (cit. on p. 17).

[49] R. P. Feynman, "Forces in Molecules", Physical Review **56**, 340–343 (1939) (cit. on p. 17).

[50] P. Pulay, "Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules", Molecular Physics **17**, 197–204 (1969) (cit. on p. 18).

[51] S. Redner, "Citation Statistics from 110 Years of Physical Review", Physics Today **58**, 49–54 (2005) (cit. on p. 19).

[52] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne, "First principles methods using CASTEP", Zeitschrift für Kristallographie - Crystalline Materials **220**, 567–570 (2005) (cit. on p. 20).

[53] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, "Using ONETEP for accurate and efficient O(N) density functional calculations", Journal of Physics: Condensed Matter **17**, 5757–5769 (2005) (cit. on pp. 20, 21, 24, 37).

[54] G. Galli, "Linear scaling methods for electronic structure calculations and quantum molecular dynamics simulations", Current Opinion in Solid State and Materials Science **1**, 864–874 (1996) (cit. on pp. 19, 87).

[55] S. Goedecker, "Linear scaling electronic structure methods", Reviews of Modern Physics **71**, 1085–1123 (1999) (cit. on p. 19).

[56] D. R. Bowler, I. J. Bush, and M. J. Gillan, "Practical methods for ab initio calculations on thousands of atoms", International Journal of Quantum Chemistry **77**, 831–842 (2000) (cit. on p. 20).

[57] J. Vandevondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach", Computer Physics Communications **167**, 103–128 (2005) (cit. on p. 20).

[58] T. Ozaki, "O(N) Krylov-subspace method for large-scale ab initio electronic structure calculations", Physical Review B **74**, 245101 (2006) (cit. on p. 20).

[59] L. Genovese, A. Neelov, S. Goedecker, T. Deutsch, S. A. Ghasemi, A. Willand, D. Caliste, O. Zilberberg, M. Rayson, A. Bergman, and R. Schneider, "Daubechies wavelets as a basis set for density functional pseudopotential calculations", The Journal of Chemical Physics **129**, 014109 (2008) (cit. on p. 20).

[60] S. Mohr, L. E. Ratcliff, L. Genovese, D. Caliste, P. Boulanger, S. Goedecker, and T. Deutsch, "Accurate and efficient linear scaling DFT calculations with universal applicability", Physical Chemistry Chemical Physics **17**, 31360–31370 (2015) (cit. on p. 20).

[61] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne, "Introducing ONETEP : Linear-scaling density functional simulations on parallel computers", The Journal of Chemical Physics **122**, 084119 (2005) (cit. on pp. 20, 37, 87, 94).

[62] W. Kohn, "Density Functional and Density Matrix Method Scaling Linearly with the Number of Atoms", Physical Review Letters **76**, 3168–3171 (1996) (cit. on p. 21).

[63] E. Prodan and W. Kohn, "Nearsightedness of electronic matter", Proceedings of the National Academy of Sciences of the United States of America **102**, 11635–8 (2005) (cit. on p. 21).

[64] S. Ismail-Beigi and T. A. Arias, "Locality of the Density Matrix in Metals, Semiconductors, and Insulators", Physical Review Letters **82**, 2127–2130 (1999) (cit. on p. 21).

[65] A. A. Mostofi, P. D. Haynes, C.-K. Skylaris, and M. C. Payne, "Preconditioned iterative minimization for linear-scaling electronic structure calculations", The Journal of Chemical Physics **119**, 8842–8848 (2003) (cit. on p. 21).

[66] R. McWeeny, "Some Recent Advances in Density Matrix Theory", Reviews of Modern Physics **32**, 335–369 (1960) (cit. on pp. 21, 23).

[67] P. D. Haynes and M. C. Payne, "Corrected penalty-functional method for linear-scaling calculations within density-functional theory", Physical Review B **59**, 12173–12176 (1999) (cit. on p. 23).

[68] X.-P. Li, R. W. Nunes, and D. Vanderbilt, "Density-matrix electronic-structure method with linear system-size scaling", Physical Review B **47**, 10891–10894 (1993) (cit. on p. 23).

[69] A. A. Mostofi, C.-K. Skylaris, P. D. Haynes, and M. C. Payne, "Total-energy calculations on a real space grid with localized functions and a plane-wave basis", Computer Physics Communications **147**, 788–802 (2002) (cit. on p. 24).

[70] C.-K. Skylaris, A. A. Mostofi, P. D. Haynes, C. J. Pickard, and M. C. Payne, "Accurate kinetic energy evaluation in electronic structure calculations with localized functions on real space grids", Computer Physics Communications **140**, 315–322 (2001) (cit. on pp. 25, 37).

[71] P. E. Blöchl, "Projector augmented-wave method", Physical Review B **50**, 17953–17979 (1994) (cit. on pp. 26, 90, 94).

[72] M. Torrent, F. Jollet, F. Bottin, G. Zérah, and X. Gonze, "Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure", Computational Materials Science **42**, 337–351 (2008) (cit. on pp. 28, 35).

[73] N. D. M. Hine, "Linear-scaling density functional theory using the projector augmented wave method", Journal of Physics: Condensed Matter **29**, 024001 (2017) (cit. on pp. 28, 87, 94).

[74] V. I. Anisimov, J. Zaanen, and O. K. Andersen, "Band theory and Mott insulators: Hubbard U instead of Stoner I", Physical Review B **44**, 943–954 (1991) (cit. on p. 29).

[75] V. I. Anisimov, I. V. Solovyev, M. A. Korotin, M. T. Czyżyk, and G. A. Sawatzky, "Density-functional theory and NiO photoemission spectra", Physical Review B **48**, 16929–16934 (1993) (cit. on p. 29).

[76] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: the LDA+U method", Journal of Physics: Condensed Matter **9**, 767–808 (1997) (cit. on p. 29).

[77] J. D. Talman and W. F. Shadwick, "Optimized effective atomic central potential", Physical Review A **14**, 36–40 (1976) (cit. on p. 29).

[78] T. Kotani, "Exact Exchange Potential Band-Structure Calculations by the Linear Muffin-Tin Orbital-Atomic-Sphere Approximation Method for Si, Ge, C, and MnO", Physical Review Letters **74**, 2989–2992 (1995) (cit. on p. 29).

[79] T. W. Hollins, S. J. Clark, K. Refson, and N. I. Gidopoulos, "Optimized effective potential using the Hylleraas variational method", Physical Review B **85**, 235126 (2012) (cit. on p. 29).

[80] L. Hedin, "New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem", Physical Review **139**, A796–A823 (1965) (cit. on pp. 30, 32).

[81] W. C. Dunlap and R. L. Watters, "Direct Measurement of the Dielectric Constants of Silicon and Germanium", Physical Review **92**, 1396–1397 (1953) (cit. on p. 30).

[82] M. van Schilfgaarde, T. Kotani, and S. Faleev, "Quasiparticle Self-Consistent GW Theory", Physical Review Letters **96**, 226402 (2006) (cit. on p. 30).

[83] F. Aryasetiawan and O. Gunnarsson, "The GW method", Reports on Progress in Physics **61**, 237–312 (1998) (cit. on pp. 30, 34).

[84] J. A. Berger, L. Reining, and F. Sottile, "Efficient GW calculations for SnO2, ZnO, and rubrene: The effective-energy technique", Physical Review B **85**, 085126 (2012) (cit. on pp. 32, 74, 75).

[85] D. Bohm and D. Pines, "A Collective Description of Electron Interactions. I. Magnetic Interactions", Physical Review **82**, 625–634 (1951) (cit. on p. 33).

[86] B. I. Lundqvist, "Single-particle spectrum of the degenerate electron gas", Physik der Kondensierten Materie **6**, 193–205 (1967) (cit. on p. 34).

[87] B. I. Lundqvist, "Single-particle spectrum of the degenerate electron gas", Physik der Kondensierten Materie **6**, 206–217 (1967) (cit. on p. 34).

[88] L. Hedin and S. Lundqvist, "Effects of Electron-Electron and Electron-Phonon Interactions on the One-Electron States of Solids", Solid State Physics - Advances in Research and Applications **23**, 1–181 (1970) (cit. on p. 34).

[89] M. S. Hybertsen and S. G. Louie, "Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies", Physical Review B **34**, 5390–5413 (1986) (cit. on p. 34).

[90] R. W. Godby and R. J. Needs, "Metal-insulator transition in Kohn-Sham theory and quasiparticle theory", Physical Review Letters **62**, 1169–1172 (1989) (cit. on p. 35).

[91] P. Larson, M. Dvorak, and Z. Wu, "Role of the plasmon-pole model in the GW approximation", Physical Review B **88**, 125205 (2013) (cit. on p. 35).

[92] J. Hafner and G. Kresse, "The Vienna AB-Initio Simulation Program VASP: An Efficient and Versatile Tool for Studying the Structural, Dynamic, and Electronic Properties of Materials", in *Properties of complex inorganic solids* (Springer US, Boston, MA, 1997), pp. 69–82 (cit. on p. 35).

[93] J. Deslippe, G. Samsonidze, D. A. Strubbe, M. Jain, M. L. Cohen, and S. G. Louie, "BerkeleyGW: A massively parallel computer package for the calculation of the quasiparticle and optical properties of materials and nanostructures", Computer Physics Communications **183**, 1269–1289 (2012) (cit. on p. 35).

[94] A. Marini, C. Hogan, M. Grüning, and D. Varsano, "yambo: An ab initio tool for excited state calculations", Computer Physics Communications **180**, 1392–1403 (2009) (cit. on pp. 35, 65, 68, 69).

[95] J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard, "Resolution of the Band Gap Prediction Problem for Materials Design", The Journal of Physical Chemistry Letters **7**, 1198–1203 (2016) (cit. on p. 35).

[96] E. E. Salpeter and H. A. Bethe, "A Relativistic Equation for Bound-State Problems", Physical Review **84**, 1232–1242 (1951) (cit. on p. 35).

[97] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, "Optical Spectrum of MoS2: Many-Body Effects and Diversity of Exciton States", Physical Review Letters **111**, 216805 (2013) (cit. on p. 35).

[98] A. Molina-Sánchez, D. Sangalli, K. Hummer, A. Marini, and L. Wirtz, "Effect of spin-orbit interaction on the optical spectra of single-layer, double-layer, and bulk MoS2", Physical Review B **88**, 045412 (2013) (cit. on p. 35).

[99] H. Mishra, A. Bose, A. Dhar, and S. Bhattacharya, "Exciton-phonon coupling and band-gap renormalization in monolayer WSe2", Physical Review B **98**, 045143 (2018) (cit. on pp. 35, 72).

[100] X.-Q. Zhang, C.-H. Lin, Y.-W. Tseng, K.-H. Huang, and Y.-H. Lee, "Synthesis of Lateral Heterostructures of Semiconducting Atomic Layers", Nano Letters **15**, 410–415 (2015) (cit. on p. 36).

[101] M.-Y. Li, Y. Shi, C.-C. Cheng, L.-S. Lu, Y.-C. Lin, H.-L. Tang, M.-L. Tsai, C.-W. Chu, K.-H. Wei, J.-H. He, W.-H. Chang, K. Suenaga, and L.-J. Li, "Epitaxial growth of a monolayer WSe2-MoS2 lateral p-n junction with an atomically sharp interface", Science **349**, 524–528 (2015) (cit. on p. 36).

[102] C. Gong, H. Zhang, W. Wang, L. Colombo, R. M. Wallace, and K. Cho, "Band alignment of two-dimensional transition metal dichalcogenides: Application in tunnel field effect transistors", Applied Physics Letters **103**, 053513 (2013) (cit. on pp. 36, 39).

[103] Y. Gong, J. Lin, X. Wang, G. Shi, S. Lei, Z. Lin, X. Zou, G. Ye, R. Vajtai, B. I. Yakobson, H. Terrones, M. Terrones, B. K. Tay, J. Lou, S. T. Pantelides, Z. Liu, W. Zhou, and P. M. Ajayan, "Vertical and in-plane heterostructures from WS2/MoS2 monolayers", Nature Materials **13**, 1135–1142 (2014) (cit. on pp. 36, 38, 39).

[104] Y. Kobayashi, S. Mori, Y. Maniwa, and Y. Miyata, "Bandgap-tunable lateral and vertical heterostructures based on monolayer Mo1-xWxS2 alloys", Nano Research **8**, 3261–3271 (2015) (cit. on p. 36).

[105] K. Chen, X. Wan, J. Wen, W. Xie, Z. Kang, X. Zeng, H. Chen, and J.-B. Xu, "Electronic Properties of MoS2-WS2 Heterostructures Synthesized with Two-Step Lateral Epitaxial Strategy", ACS Nano **9**, 9868–9876 (2015) (cit. on p. 36).

[106] K. F. Garrity, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, "Pseudopotentials for high-throughput DFT calculations", Computational Materials Science **81**, 446–452 (2014) (cit. on p. 37).

[107] W. J. Schutte, J. L. De Boer, and F. Jellinek, "Crystal structures of tungsten disulfide and diselenide", Journal of Solid State Chemistry **70**, 207–209 (1987) (cit. on p. 37).

[108] V. Petkov, S. J. L. Billinge, P. Larson, S. D. Mahanti, T. Vogt, K. K. Rangan, and M. G. Kanatzidis, "Structure of nanocrystalline materials using atomic pair distribution function analysis: Study of LiMoS2", Physical Review B **65**, 092105 (2002) (cit. on pp. 37, 41, 51).

[109] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, "Atomically Thin MoS2: A New Direct-Gap Semiconductor", Physical Review Letters **105**, 136805 (2010) (cit. on p. 38).

[110] N. D. M. Hine, P. W. Avraam, P. Tangney, and P. D. Haynes, "Linear-scaling density functional theory simulations of polar semiconductor nanorods", Journal of Physics: Conference Series **367**, 012002 (2012) (cit. on p. 38).

[111] J. Aarons, L. G. Verga, N. D. M. Hine, and C.-K. Skylaris, "Atom-projected and angular momentum resolved density of states in the ONETEP code", Electronic Structure **1**, 035002 (2019) (cit. on p. 38).

[112] J. Kang, S. Tongay, J. Zhou, J. Li, and J. Wu, "Band offsets and heterostructures of two-dimensional semiconductors", Applied Physics Letters **102**, 012111 (2013) (cit. on p. 39).

[113] S. Zhang and J. Northrup, "Chemical potential dependence of defect formation energies in GaAs: Application to Ga self-diffusion", Physical Review Letters **67**, 2339–2342 (1991) (cit. on p. 39).

[114] Z. Lin, B. R. Carvalho, E. Kahn, R. Lv, R. Rao, H. Terrones, M. A. Pimenta, and M. Terrones, "Defect engineering of two-dimensional transition metal dichalcogenides", 2D Materials **3**, 022002 (2016) (cit. on p. 41).

[115] K. Lejaeghere, V. Van Speybroeck, G. Van Oost, and S. Cottenier, "Error Estimates for Solid-State Density-Functional Theory Predictions: An Overview by Means of the Ground-State Elemental Crystals", Critical Reviews in Solid State and Materials Sciences **39**, 1–24 (2014) (cit. on p. 41).

[116] B. C. McGuigan, P. Pochet, and H. T. Johnson, "Critical thickness for interface misfit dislocation formation in two-dimensional materials", Physical Review B **93**, 214103 (2016) (cit. on pp. 45, 50, 51).

[117] B. I. Yakobson, "Mechanical relaxation and "intramolecular plasticity" in carbon nanotubes", Applied Physics Letters **72**, 918–920 (1998) (cit. on p. 45).

[118] X. Zou, Y. Liu, and B. I. Yakobson, "Predicting Dislocations and Grain Boundaries in Two-Dimensional Metal-Disulfides from the First Principles", Nano Letters **13**, 253–258 (2013) (cit. on p. 46).

[119] W. Zhou, X. Zou, S. Najmaei, Z. Liu, Y. Shi, J. Kong, J. Lou, P. M. Ajayan, B. I. Yakobson, and J.-C. Idrobo, "Intrinsic Structural Defects in Monolayer Molybdenum Disulfide", Nano Letters **13**, 2615–2622 (2013) (cit. on p. 46).

[120] J. M. Burgers, "Geometrical considerations concerning the structural irregularities to be assumed in a crystal", Proceedings of the Physical Society **52**, 23–33 (1940) (cit. on p. 47).

[121] P. B. James, M. T. Lavik, and IUCr, "The crystal structure of MoSe2", Acta Crystallographica **16**, 1183–1183 (1963) (cit. on p. 51).

[122] P. Dudin, P. Lacovig, C. Fava, E. Nicolini, A. Bianco, G. Cautero, and A. Barinov, "Angle-resolved photoemission spectroscopy and imaging with a submicrometre probe at the Spectromicroscopy-3.2L beamline of Elettra", Journal of Synchrotron Radiation **17**, 445–450 (2010) (cit. on pp. 58, 61).

[123] E. Rotenberg and A. Bostwick, "microARPES and nanoARPES at diffraction-limited light sources: opportunities and performance gains", Journal of Synchrotron Radiation **21**, 1048–1056 (2014) (cit. on p. 58).

[124] H. Iwasawa, P. Dudin, K. Inui, T. Masui, T. K. Kim, C. Cacho, and M. Hoesch, "Buried double CuO chains in YBa2Cu4O8 uncovered by nano-ARPES", Physical Review B **99**, 140510 (2019) (cit. on p. 58).

[125] T. Sato and T. Takahashi, "Angle-Resolved Photoemission Spectroscopy of Graphene, Graphite, and Related Compounds", in *Comprehensive semiconductor science and technology* (Elsevier, 2011), pp. 383–409 (cit. on p. 58).

[126] R. J. Koch, C. Jozwiak, A. Bostwick, B. Stripe, M. Cordier, Z. Hussain, W. Yun, and E. Rotenberg, "Nano focusing of soft X-rays by a new capillary mirror optic", Synchrotron Radiation News **31**, 50–52 (2018) (cit. on p. 58).

[127] H. Zhang, C. Bao, Z. Jiang, K. Zhang, H. Li, C. Chen, J. Avila, Y. Wu, W. Duan, M. C. Asensio, and S. Zhou, "Resolving Deep Quantum-Well States in Atomically Thin 2H-MoTe2 Flakes by Nanospot Angle-Resolved Photoemission Spectroscopy", Nano Letters **18**, 4664–4668 (2018) (cit. on pp. 58, 60).

[128] K. S. Novoselov, A. Mishchenko, A. Carvalho, and A. H. Castro Neto, "2D materials and van der Waals heterostructures.", Science (New York, N.Y.) **353**, aac9439 (2016) (cit. on p. 58).

[129] J. Katoch, S. Ulstrup, R. J. Koch, S. Moser, K. M. McCreary, S. Singh, J. Xu, B. T. Jonker, R. K. Kawakami, A. Bostwick, E. Rotenberg, and C. Jozwiak, "Giant spin-splitting and gap renormalization driven by trions in single-layer WS2/h-BN heterostructures", Nature Physics **14**, 355–359 (2018) (cit. on pp. 58, 60).

[130] I. Cucchi, I. Gutiérrez-Lezama, E. Cappelli, S. McKeown Walker, F. Y. Bruno, G. Tenasini, L. Wang, N. Ubrig, C. Barreteau, E. Giannini, M. Gibertini, A. Tamai, A. F. Morpurgo, and F. Baumberger, "Microfocus Laser-Angle-Resolved Photoemission on Encapsulated Mono-, Bi-, and Few-Layer 1T'-WTe2", Nano Letters **19**, 554–560 (2019) (cit. on p. 58).

[131] N. R. Wilson, P. V. Nguyen, K. Seyler, P. Rivera, A. J. Marsden, Z. P. Laker, G. C. Constantinescu, V. Kandyba, A. Barinov, N. D. Hine, X. Xu, and D. H. Cobden, "Determination of band offsets, hybridization, and exciton binding in 2D semiconductor heterostructures", Science Advances **3**, e1601832 (2017) (cit. on pp. 58, 85).

[132] W. Jin, P.-C. Yeh, N. Zaki, D. Chenet, G. Arefe, Y. Hao, A. Sala, T. O. Mentes, J. I. Dadap, A. Locatelli, J. Hone, and R. M. Osgood, "Tuning the electronic structure of monolayer graphene/MoS2 van der Waals heterostructures via interlayer twist", Physical Review B **92**, 201409 (2015) (cit. on p. 58).

[133] D. Pierucci, H. Henck, J. Avila, A. Balan, C. H. Naylor, G. Patriarche, Y. J. Dappe, M. G. Silly, F. Sirotti, A. T. C. Johnson, M. C. Asensio, and A. Ouerghi, "Band Alignment and Minigaps in Monolayer MoS2-Graphene van der Waals Heterostructures", Nano Letters **16**, 4054–4061 (2016) (cit. on p. 58).

[134] B. O. Community, *Blender - a 3D modelling and rendering package*, Stichting Blender Foundation, Amsterdam, 2018 (cit. on p. 59).

[135] J. M. Riley, W. Meevasana, L. Bawden, M. Asakawa, T. Takayama, T. Eknapakul, T. K. Kim, M. Hoesch, S.-K. Mo, H. Takagi, T. Sasagawa, M. S. Bahramy, and P. D. C. King, "Negative electronic compressibility and tunable spin splitting in WSe2", Nature Nanotechnology **10**, 1043–1047 (2015) (cit. on p. 60).

[136] Y. Zhang, M. M. Ugeda, C. Jin, S.-F. Shi, A. J. Bradley, A. Martín-Recio, H. Ryu, J. Kim, S. Tang, Y. Kim, B. Zhou, C. Hwang, Y. Chen, F. Wang, M. F. Crommie, Z. Hussain, Z.-X. Shen, and S.-K. Mo, "Electronic Structure, Surface Doping, and Optical Response in Epitaxial WSe2 Thin Films", Nano Letters **16**, 2485–2491 (2016) (cit. on p. 60).

[137] A. Bostwick, T. Ohta, T. Seyller, K. Horn, and E. Rotenberg, "Quasiparticle dynamics in graphene", Nature Physics **3**, 36–40 (2007) (cit. on p. 60).

[138] M. Yankowitz, D. McKenzie, and B. J. LeRoy, "Local Spectroscopic Characterization of Spin and Layer Polarization in WSe2", Physical Review Letters **115**, 136803 (2015) (cit. on p. 60).

[139] C. Zhang, Y. Chen, A. Johnson, M.-Y. Li, L.-J. Li, P. C. Mende, R. M. Feenstra, and C.-K. Shih, "Probing Critical Point Energies of Transition Metal Dichalcogenides: Surprising Indirect Gap of Single Layer WSe2", Nano Letters **15**, 6494–6500 (2015) (cit. on p. 60).

[140] K. He, N. Kumar, L. Zhao, Z. Wang, K. F. Mak, H. Zhao, and J. Shan, "Tightly Bound Excitons in Monolayer WSe2", Physical Review Letters **113**, 026803 (2014) (cit. on pp. 60, 72).

[141] S. Park, N. Mutz, T. Schultz, S. Blumstengel, A. Han, A. Aljarb, L.-J. Li, E. J. W. List-Kratochvil, P. Amsalem, and N. Koch, "Direct determination of monolayer MoS2 and WSe2 exciton binding energies on insulating and metallic substrates", 2D Materials **5**, 025003 (2018) (cit. on p. 60).

[142] A. V. Stier, N. P. Wilson, K. A. Velizhanin, J. Kono, X. Xu, and S. A. Crooker, "Magnetooptics of Exciton Rydberg States in a Monolayer Semiconductor", Physical Review Letters **120**, 057405 (2018) (cit. on p. 60).

[143] W. Zhao, R. M. Ribeiro, and G. Eda, "Electronic Structure and Optical Signatures of Semiconducting Transition Metal Dichalcogenide Nanosheets", Accounts of Chemical Research **48**, 91–99 (2015) (cit. on pp. 60, 62).

[144] P. J. Zomer, M. H. D. Guimarães, J. C. Brant, N. Tombros, and B. J. van Wees, "Fast pick up technique for high quality heterostructures of bilayer graphene and hexagonal boron nitride", Applied Physics Letters **105**, 013101 (2014) (cit. on p. 60).

[145] S. Luryi, "Quantum capacitance devices", Applied Physics Letters **52**, 501–503 (1988) (cit. on p. 61).

[146] T. Brumme, M. Calandra, and F. Mauri, "First-principles theory of field-effect doping in transition-metal dichalcogenides: Structural properties, electronic structure, Hall coefficient, and electrical conductivity", Physical Review B **91**, 155436 (2015) (cit. on pp. 61, 66).

[147] G.-B. Liu, D. Xiao, Y. Yao, X. Xu, and W. Yao, "Electronic structures and theoretical modelling of two-dimensional group-VIB transition metal dichalcogenides", Chemical Society Reviews **44**, 2643–2663 (2015) (cit. on p. 62).

[148] D. R. Hamann, "Optimized norm-conserving Vanderbilt pseudopotentials", Physical Review B **88**, 085117 (2013) (cit. on pp. 65, 67, 77).

[149] M. van Setten, M. Giantomassi, E. Bousquet, M. Verstraete, D. Hamann, X. Gonze, and G.-M. Rignanese, "The PseudoDojo: Training and grading a 85 element optimized norm-conserving pseudopotential table", Computer Physics Communications **226**, 39–54 (2018) (cit. on pp. 65, 67, 77).

[150] S. Gao and L. Yang, "Renormalization of the quasiparticle band gap in doped two-dimensional materials from many-body calculations", Physical Review B **96**, 155410 (2017) (cit. on p. 66).

[151] A. Bruix, J. A. Miwa, N. Hauptmann, D. Wegner, S. Ulstrup, S. S. Grønborg, C. E. Sanders, M. Dendzik, A. Grubišić Čabo, M. Bianchi, J. V. Lauritsen, A. A. Khajetoorians, B. Hammer, and P. Hofmann, "Single-layer MoS2 on Au(111): Band gap renormalization and substrate interaction", Physical Review B **93**, 165422 (2016) (cit. on p. 66).

[152] M. Rohlfing, P. Krüger, and J. Pollmann, "Quasiparticle Band Structure of CdS", Physical Review Letters **75**, 3489–3492 (1995) (cit. on p. 67).

[153] M. Rohlfing, P. Krüger, and J. Pollmann, "Quasiparticle calculations of semicore states in Si, Ge, and CdS", Physical Review B **56**, R7065–R7068 (1997) (cit. on p. 67).

[154] M. Rohlfing, P. Krüger, and J. Pollmann, "Role of semicore d electrons in quasiparticle band-structure calculations", Physical Review B **57**, 6485–6492 (1998) (cit. on p. 67).

[155] A. Marini, R. Del Sole, A. Rubio, and G. Onida, "Quasiparticle band-structure effects on the d hole lifetimes of copper within the GW approximation", Physical Review B **66**, 161104 (2002) (cit. on p. 67).

[156] T. Rangel, D. Kecik, P. E. Trevisanutto, G.-M. Rignanese, H. Van Swygenhoven, and V. Olevano, "Band structure of gold from many-body perturbation theory", Physical Review B **86**, 125125 (2012) (cit. on p. 67).

[157] D. Golze, M. Dvorak, and P. Rinke, "The GW Compendium: A Practical Guide to Theoretical Photoemission Spectroscopy", Frontiers in Chemistry **7** (2019) (cit. on p. 67).

[158] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, S. de Gironcoli, T. Deutsch, J. K. Dewhurst, I. Di Marco, C. Draxl, M. Dułak, O. Eriksson, J. A. Flores-Livas, K. F. Garrity, L. Genovese, P. Giannozzi, M. Giantomassi, S. Goedecker, X. Gonze, O. Grånäs, E. K. U. Gross, A. Gulans, F. Gygi, D. R. Hamann, P. J. Hasnip, N. A. W. Holzwarth, D. Iuşan, D. B. Jochym, F. Jollet, D. Jones, G. Kresse, K. Koepernik, E. Küçükbenli, Y. O. Kvashnin, I. L. M. Locht, S. Lubeck, M. Marsman, N. Marzari, U. Nitzsche, L. Nordström, T. Ozaki, L. Paulatto, C. J. Pickard, W. Poelmans, M. I. J. Probert, K. Refson, M. Richter, G.-M. Rignanese, S. Saha, M. Scheffler, M. Schlipf, K. Schwarz, S. Sharma, F. Tavazza, P. Thunström, A. Tkatchenko, M. Torrent, D. Vanderbilt, M. J. van Setten, V. Van Speybroeck, J. M. Wills, J. R. Yates, G.-X. Zhang, and S. Cottenier, "Reproducibility in density functional theory

calculations of solids.", Science (New York, N.Y.) **351**, aad3000 (2016) (cit. on p. 67).

[159] S. Haastrup, M. Strange, M. Pandey, T. Deilmann, P. S. Schmidt, N. F. Hinsche, M. N. Gjerding, D. Torelli, P. M. Larsen, A. C. Riis-Jensen, J. Gath, K. W. Jacobsen, J. Jørgen Mortensen, T. Olsen, and K. S. Thygesen, "The Computational 2D Materials Database: high-throughput modeling and discovery of atomically thin crystals", 2D Materials **5**, 042002 (2018) (cit. on pp. 72, 85, 94, 99).

[160] W. Wang, A. Narayan, L. Tang, K. Dolui, Y. Liu, X. Yuan, Y. Jin, Y. Wu, I. Rungger, S. Sanvito, and F. Xiu, "Spin-Valve Effect in NiFe/MoS2/NiFe Junctions", Nano Letters **15**, 5261–5267 (2015) (cit. on p. 72).

[161] B.-C. Shih, Y. Xue, P. Zhang, M. L. Cohen, and S. G. Louie, "Quasiparticle Band Gap of ZnO: High Accuracy from the Conventional G0W0 Approach", Physical Review Letters **105**, 146401 (2010) (cit. on p. 74).

[162] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, "Screening and many-body effects in two-dimensional crystals: Monolayer MoS2", Physical Review B **93**, 235435 (2016) (cit. on p. 74).

[163] J.-G. Song, G. H. Ryu, S. J. Lee, S. Sim, C. W. Lee, T. Choi, H. Jung, Y. Kim, Z. Lee, J.-M. Myoung, C. Dussarrat, C. Lansalot-Matras, J. Park, H. Choi, and H. Kim, "Controllable synthesis of molybdenum tungsten disulfide alloy for vertically composition-controlled multilayer", Nature Communications **6**, 7817 (2015) (cit. on p. 79).

[164] Z. Wang, P. Liu, Y. Ito, S. Ning, Y. Tan, T. Fujita, A. Hirata, and M. Chen, "Chemical Vapor Deposition of Monolayer Mo1-xWxS2 Crystals with Tunable Band Gaps", Scientific Reports **6**, 21536 (2016) (cit. on p. 79).

[165] N. J. Ramer and A. M. Rappe, "Application of a new virtual crystal approach for the study of disordered perovskites", Journal of Physics and Chemistry of Solids **61**, 315–320 (2000) (cit. on p. 79).

[166] Y. Chen, J. Xi, D. O. Dumcenco, Z. Liu, K. Suenaga, D. Wang, Z. Shuai, Y.-S. Huang, and L. Xie, "Tunable Band Gap Photoluminescence from Atomically Thin Transition-Metal Dichalcogenide Alloys", ACS Nano **7**, 4610–4616 (2013) (cit. on p. 79).

[167] T. Georgiou, R. Jalil, B. D. Belle, L. Britnell, R. V. Gorbachev, S. V. Morozov, Y.-J. Kim, A. Gholinia, S. J. Haigh, O. Makarovsky, L. Eaves, L. A. Ponomarenko, A. K. Geim, K. S. Novoselov, and A. Mishchenko, "Vertical field-effect transistor based on graphene-WS2 heterostructures for flexible and transparent electronics", Nature Nanotechnology **8**, 100–103 (2013) (cit. on p. 84).

[168] C.-J. Shih, Q. H. Wang, Y. Son, Z. Jin, D. Blankschtein, and M. S. Strano, "Tuning On-Off Current Ratio and Field-Effect Mobility in a MoS2-Graphene Heterostructure via Schottky Barrier Modulation", ACS Nano **8**, 5790–5798 (2014) (cit. on p. 84).

[169] L. Britnell, R. M. Ribeiro, A. Eckmann, R. Jalil, B. D. Belle, A. Mishchenko, Y.-J. Kim, R. V. Gorbachev, T. Georgiou, S. V. Morozov, A. N. Grigorenko, A. K. Geim, C. Casiraghi, A. H. Castro Neto, and K. S. Novoselov, "Strong light-matter interactions in heterostructures of atomically thin films.", Science (New York, N.Y.) **340**, 1311–4 (2013) (cit. on p. 84).

[170] M. Bernardi, M. Palummo, and J. C. Grossman, "Extraordinary Sunlight Absorption and One Nanometer Thick Photovoltaics Using Two-Dimensional Monolayer Materials", Nano Letters **13**, 3664–3670 (2013) (cit. on p. 84).

[171] M. M. Furchi, A. Pospischil, F. Libisch, J. Burgdörfer, and T. Mueller, "Photovoltaic Effect in an Electrically Tunable van der Waals Heterojunction", Nano Letters **14**, 4785–4791 (2014) (cit. on p. 84).

[172] J. Kang, J. Li, S.-S. Li, J.-B. Xia, and L.-W. Wang, "Electronic Structural Moiré Pattern Effects on MoS2/MoSe 2 2D Heterostructures", Nano Letters **13**, 5485–5490 (2013) (cit. on p. 85).

[173] G. C. Constantinescu and N. D. M. Hine, "Energy landscape and band-structure tuning in realistic MoS2/MoSe2 heterostructures", Physical Review B **91**, 195416 (2015) (cit. on pp. 85, 94, 97, 102).

[174] L. E. Ratcliff, N. D. M. Hine, and P. D. Haynes, "Calculating optical absorption spectra for large systems using linear-scaling density functional theory", Physical Review B **84**, 165131 (2011) (cit. on p. 85).

[175] F. Jollet, M. Torrent, and N. Holzwarth, "Generation of Projector Augmented-Wave atomic data: A 71 element validated table in the XML format", Computer Physics Communications **185**, 1246–1254 (2014) (cit. on pp. 90, 94).

[176] W. Huang, X. Luo, C. K. Gan, S. Y. Quek, and G. Liang, "Theoretical study of thermoelectric properties of few-layer MoS2 and WSe2", Physical Chemistry Chemical Physics **16**, 10866 (2014) (cit. on pp. 94, 103).

[177] Á. Ruiz-Serrano and C.-K. Skylaris, "A variational method for density functional theory calculations on metallic systems with thousands of atoms", The Journal of Chemical Physics **139**, 054107 (2013) (cit. on p. 97).

[178] S. Gao, L. Yang, and C. D. Spataru, "Interlayer Coupling and Gate-Tunable Excitons in Transition Metal Dichalcogenide Heterostructures", Nano Letters **17**, 7809–7813 (2017) (cit. on p. 97).